

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 15, 2003, 18:55:03 ; Search time 5026 Seconds  
(without alignments)  
10711.713 Million cell updates/sec

Title: US-10-033-269-8  
Perfect score: 1316  
Sequence: 1 tcggcacgaggctttctcag.....tggcattgctatttgctagg 1316

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:				
1:	gb_ba:			A94667 Sequence 5
2:	gb_htg:			Y10252 L.japonicus
3:	gb_in:			A94671 Sequence 9
4:	gb_om:			AY085534 Arabidops
5:	gb_ov:			A94668 Sequence 6
6:	gb_pat:			Y10253 O.sativa pa
7:	gb_ph:			AB012242 Arabidops
8:	gb_pl:			AC130200 Medicago
9:	gb_pr:			AC096688 Oryza sat
10:	gb_ro:			AJ007446 Thermotog
11:	gb_sts:			AE001768 Thermotog
12:	gb_sy:			BX321856 Nitrosomo
13:	gb_un:			AR204111 Sequence
14:	gb_vi:			AE005189 Escherich
15:	em_ba:			AP002550 Escherich
16:	em_fun:			Z49811 S.pombe chr
17:	em_hum:			AE016791 Pseudomon
18:	em_in:			AE016755 Escherich
19:	em_mu:			AE016952 Enterococ
20:	em_om:			AE017002 Bacillus
21:	em_or:			AY142936 Heliobaci
22:	em_ov:			AE007789 Clostridi
23:	em_ph:			AE015532 Shewanell
24:	em_pl:			AE011810 Xanthomon
25:	em_ro:			LI7086 E. coli alp
26:	em_sts:			AE000122 Escherich
27:	em_un:			AX370251 Sequence
28:	em_vi:			AP004830 Staphyloc
29:	em_htg_hum:			AE017029 Bacillus
30:	em_htg_inv:			AX143131 Sequence
31:	em_htg_other:			AF270140 Staphyloc
32:	em_htg_mus:			AX145458 sequence
33:	em_htg_pln:			AE016751 Staphyloc
34:	em_htg_rod:			AE015050 Shigella
35:	em_htg_mam:			AE016978 Shigella
36:	em_htg_vrt:			AP003137 Staphyloc
37:	em_sy:			AP003365 Staphyloc
38:	em_htgo_hum:			AX620824 Sequence
39:	em_htgo_mus:			AC074026 Staphyloc
40:	em_htgo_other:			AJ298881 Fusarium
41:				AE003876 Xylella f
				AL139074 Campyloba
				AL162755 Neisseria
				AE012553 Xylella f
				AE004886 Pseudomon

ALIGNMENTS

RESULT 1

A94667

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

A94667  
Sequence 5 from Patent WO9942565.

A94667

A94667.1

GI:6778943

unidentified

unclassified

1 (bases 1 to 1338)

Laber,B. and Genschel,U.

HERBICIDES TEST METHOD

Patent: WO 9942565-A 5 26-AUG-1999;

LABER BERND (DE); GENSCHER ULRICH (DE)

DNA

1338 bp

linear

PAT 26-JAN-2000

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	625	47.5	1338	6	A94667	A94667 Sequence 5
2	625	47.5	1338	8	LJPANC	Y10252 L.japonicus
3	599.6	45.6	1347	6	A94671	A94671 Sequence 9
4	422.2	32.1	1188	8	AY085534	AY085534 Arabidops
5	354	26.9	1264	6	A94668	A94668 Sequence 6
6	354	26.9	1264	8	OSPANC	Y10253 O.sativa pa
7	256	19.5	78973	8	AB012242	AB012242 Arabidops
8	246.4	18.7	118998	2	AC130200	AC130200 Medicago
9	226	17.2	145290	8	AC096688	AC096688 Oryza sat
10	147.6	11.2	4283	1	TNE7446	AJ007446 Thermotog
11	135.6	10.3	19711	1	AE001768	AE001768 Thermotog
12	131.6	10.0	302050	1	BX321856	BX321856 Nitrosomo
13	118.8	9.0	9381	6	AR204111	AR204111 Sequence
14	118.8	9.0	11214	1	AE005189	AE005189 Escherich
15	118.8	9.0	281530	1	AP002550	AP002550 Escherich
16	118.4	9.0	24845	8	SPAC5H10	Z49811 S.pombe chr
17	113.8	8.6	300933	1	AE016791	AE016791 Pseudomon
18	110.8	8.4	300409	1	AE016755	AE016755 Escherich
19	108	8.2	300029	1	AE016952	AE016952 Enterococ
20	107.6	8.2	304680	1	AE017002	AE017002 Bacillus
21	105.6	8.0	945	1	AY142936	AY142936 Heliobaci
22	105.6	8.0	11662	1	AE007789	AE007789 Clostridi
23	101.8	7.7	10709	1	AE015532	AE015532 Shewanell
24	101.6	7.7	10880	1	AE011810	AE011810 Xanthomon
25	101.2	7.7	4589	1	ECOPANBCD	LI7086 E. coli alp
26	101.2	7.7	10982	1	AE000122	AE000122 Escherich
27	101.2	7.7	10982	6	AX370251	AX370251 Sequence
28	101.2	7.7	301250	1	AP004830	AP004830 Staphyloc
29	100.2	7.6	290525	1	AE017029	AE017029 Bacillus
30	99.8	7.6	690	6	AX143131	AX143131 Sequence
31	99.8	7.6	4103	1	AF270140	AF270140 Staphyloc
32	99.8	7.6	4103	6	AX145458	AX145458 sequence
33	99.8	7.6	300275	1	AE016751	AE016751 Staphyloc
34	99.6	7.6	13180	1	AE015050	AE015050 Shigella
35	99.6	7.6	290029	1	AE016978	AE016978 Shigella
36	99.6	7.6	297850	1	AP003137	AP003137 Staphyloc
37	99.6	7.6	341350	1	AP003365	AP003365 Staphyloc
38	98	7.4	864	6	AX620824	AX620824 Sequence
39	98	7.4	52695	1	AC074026	AC074026 Staphyloc
40	97.6	7.4	1244	8	FOX298881	AJ298881 Fusarium
41	96.6	7.3	11217	1	AE003876	AE003876 Xylella f
42	96.4	7.3	314150	1	CJ11168X1	AL139074 Campyloba
43	96.4	7.3	331801	1	NMA422491	AL162755 Neisseria
44	95	7.2	300029	1	AE012553	AE012553 Xylella f
45	93.4	7.1	10929	1	AE004886	AE004886 Pseudomon

FEATURES		Location/Qualifiers			
source		1..1338			
		/organism="unidentified"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:32644"			
exon		1..1338			
BASE COUNT		380 a	242 c	328 g	388 t
ORIGIN					
	Query Match	47.5%; Score 625; DB 6; Length 1338;			
	Best Local Similarity	79.0%; Pred. No. 1.7e-110;			
	Matches 800; Conservative	0;	Mismatches 195;	Indels 18;	Gaps 4;
QY	194	GCCCCAAGGTGATCTCCGACAAGGCCTCGATGCGGAGCTGGTCGCGCTCGATCGGGGCC	253		
Db	23	GCACCAATGGTGATATCTGATAAGGACGAGATGCGGAAATGGTCAAGGTCCATGCGATCC	82		
QY	254	CAGGGCAAGCTCATTTGGGCTGGTCCCCACCATGGGCTTCCTCCACGCGGGCCACCTCTCG	313		
Db	83	CAAGGCAAGCTCATCGCCCTCGTTCCCAACCATGGGCTTCCTTCACGAAGGCCACCTTCT	142		
QY	314	CTCGTGGCCCGAGGCCCGCAACTCTCCGACGTCGTGCGCCGCTCTCCATCTACGTCAACCG	373		
Db	143	CTCGTCAGACGCTCACAAACCACGCTGACCTCGTCGCCGCTCTCAATCTATGTCAACCT	202		
QY	374	GGCCAATTCCGCCCCACGGAGGACCTCTCCACCTACCCCTCCGACTTCGACGGCGGACGTA	433		
Db	203	GGCCAGTTTTCCCGACCGAGGACCTTTCGCGCATACCTTCTGATTTTCAAGGTGATCTC	262		
QY	434	AAGAAACTCGGTCGTTCCCGGGCGGCTCGACGTCGTTTTCCATCCCCGTAACTTGTAC	493		
Db	263	CAAAACTCATGCTGTTCCCTGGTGGTGTGATGTTGTTTTCCACCCCCACAATTTGTAT	322		
QY	494	GATTACGGGAAGAACGGTGGTGGTGACGTGGCAGAG-----GCTGGTGGAAATGGTGTG	547		
Db	323	GATTACGTGGTGATGGCGGTGATGCTGTGGCGGAGTGTGGTGGTGTGGGTGGGTGCT	382		
QY	548	TGCGTTGA-----GAGTGGGTCCGGGCACGAAAGTTGGTGAGGGTTGAGAAGCTGGAA	601		
Db	383	TGTGTTGATAGGAGGAGTGGTTTTTGGGCATGAAACTTGGTTAGAGCTGAGAAGCTGGAG	442		
QY	602	TTGGGCTGTGTGGGAAGAGCAGGCCCGTTTTCTTCAGAGGGGTGGCGACTGTGGTGACG	661		
Db	443	AAACCCCTTTGTGGGAAGAGTAGGCCTGTTTTCTTTAGAGGGGTGCCACCATTTGTACC	502		
QY	662	AAGTTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAG	721		
Db	503	AAGTTGTTTAATATTGTGGAGCCTGATGTTGCTGTGTTTGGGAAGAAGGACTATCAGCAA	562		
QY	722	TGGCGGCTTAATTCAGAGGATGGTTCGAGATCTTGATTTTTCCATAAAAGTGATAGTGCT	781		
Db	563	TGGAAAAATTATTCAGAGAATGGTTCGAGATCTTGATTTTTCCATAAAAGTGATAGGTCT	622		
QY	782	GAATAACACGTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTCACCT	841		
Db	623	GAAGTAATACGTGAGAAAGATGGCCTAGCAATGAGTTCGCGTAATGTGTACCTATCACCT	682		
QY	842	GAAGAGAGGGAAAAGGCACCTATCAATAATAAATCAATTGTTAAGAGCTAAATCAGCAGCA	901		
Db	683	GAAGAGAGGGAAAAGGCAGTATCTATAAATAAATCAATTGTTTAGAGCTAAATCGGCAGCA	742		
QY	902	GGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTGTACTGAT	961		
Db	743	GAAGATGGACAGATACATTGTGAGAAATTGATAAACTTGGTCGTGCAAAAGTATCACCGAA	802		
QY	962	GCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGAA	1021		
Db	803	GCTGGTGAAGGATTGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGAA	862		
QY	1022	CAGATCAAGAGTCCTGTCGTCCTTCTGTGTTGCTGCATGGTTTTGGCAAAGTCAGGCTTATA	1081		
Db	863	TGGATCAAGGGTCCTGTTGTCTTCTGTGTTTCTGCATGGTTTTGGAAAGCCAGGCTTATA	922		

QY	1082	GACAACATGGAAATCAACTTGTCAATGAATGT- - - -TTGATCTAACCTTCTG--TCATCT	1135	
Db	923	GACAACATAGAAATCAACTTGTAAATGGAAGTAAGATTGATCTAACCTTGTGAATAATCT	982	
QY	1136	CAACATGGGCCACATGCTTAATTAAATAGTTCGGGCCACGTCCTTAACAATTC	1188	
Db	983	CAGACATGGACCATATGATTAGTTCTTGGCATTCATGGGTATAGACTTC	1035	
RESULT 2				
LJPANC				
LOCUS				
DEFINITION L.japonicus panC gene.				
ACCESSION Y10252				
VERSION Y10252.1 GI:2292920				
KEYWORDS panC gene; pantoate-beta-alanine ligase.				
SOURCE Lotus japonicus				
ORGANISM Lotus japonicus				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;				
Lotus.				
1				
Genschel,U., Powell,C.A., Abell,C. and Smith,A.G.				
The final step of pantothenate biosynthesis in higher plants:				
cloning and characterization of pantothenate synthetase from Lotus				
japonicus and Oryza sativum (rice)				
Biochem. J. 341 (Pt 3), 669-678 (1999)				
JOURNAL				
MEDLINE 99348031				
PUBMED 10417331				
REFERENCE 2 (bases 1 to 1338)				
AUTHORS Genschel,U.				
TITLE Direct Submission				
JOURNAL Submitted (23-DEC-1996) U. Genschel, University Of Cambridge,				
Department Of Plant Sciences, Downing Street, Cambridge, CB2 3EA,				
UK				
FEATURES				
source				
1..1338				
/organism="Lotus japonicus"				
/mol_type="mRNA"				
/sub_species="(Regel) Larsen"				
/db_xref="taxon:34305"				
/tissue_type="root nodules"				
/dev_stage="3 weeks"				
1..1338				
/gene="panC"				
20..946				
/gene="panC"				
/EC_number="6.3.2.1"				
/codon_start=1				
/evidence=experimental				
/product="pantoate--beta-alanine ligase"				
/protein_id="CAA71302.1"				
/db_xref="GI:2292921"				
/db_xref="SPTREMBL:O24035"				
/translation="MAPMVISDKDEMRKMSRSMRSQKLIALVPTMGFLHEGHLSLVR"				
DAHNHADLVAVSIYVNPQFSPTEDLSAYPSDFQDLQKLSVPGVDVVFHPHNLVD				
YGGDGDVAECCGDGVVSCVDRRSFGFHETWVRAEKLEKPLCGKSRPVFFRGGVATIV				
TKLFNIVEPDVAVFGKDYQQWKIIQRMVRDLDFSIKVIGSEVIREKDLAMSSRNIV				
LSPEEREKAVSINKSLFRAKSAEDGQIHCCKLINLVQSITEAGGRIDYAEIVDQNN				
LEKVEWIKGPVVFVCSAWFGKARLDNIEINL"				
BASE COUNT 380 a 242 c 328 g 388 t				
ORIGIN				
Query Match 47.5%; Score 625; DB 8; Length 1338;				
Best Local Similarity 79.0%; Pred. No. 1.7e-110;				
Matches 800; Conservative 0; Mismatches 195; Indels 18; Gaps 4;				
QY	194	GCCCCAAGGTGATCTCGACAAAGGCCCTCGATCGGAGCTGGTCGCGCTCGATCGGGCC	253	
Db	23	GCACCAATGGTATATCTGATAAGGACGAGATCGGAAATGGTCAAGTCCATGCGATCC	82	
QY	254	CAGGCAAGCTCATTTGGCTGGTCCCAACCATGGGCTTCTCCACGGCGGCCACCTCTCG	313	

Db	83	CAAGGCAAGCTCATCGCCCTCGTTCCACCATGGGCTTCCCTTACGAAGGCCACCTTTCT	142
QY	314	CTCGTGGCCCGCCGCGCAACTCTCCGACGTCGTGCGCCGTCTCCATCTACGTCAACCCG	373
Db	143	CTCGTCAGAGACGCTCACAAACGCTGACCTCGTCGCCGCTCTCAATCTATGTCAACCCCT	202
QY	374	GGCCAAATTCGCCCCCACGGAGACCTCTCCACCTACCCCTCCGACTTCGACGGCGACGTA	433
Db	203	GGCCAGTTTTCCCGACCGAGGACCTTCCGCATACCCCTTCTGATTTTCAAGGTGATCTC	262
QY	434	AAGAAACTCGCGTCCGTTCCCGCGCGCTCGACGTCGTTTTCCATCCCCGTAACTTGTAC	493
Db	263	CAAAAACCTCATGTCTGTTCTGCTGGTGTGATGTTGTTTTCCACCCCCACAAATTTGTAT	322
QY	494	GATTACGGGAAGAACGTTGGTGGTGACGTGGCAGAG-----GCTGGTGAATGGTGTG	547
Db	323	GATTACGGTGGTGATGGCGGTGATGCTGTGGCGGAGTGTGGTGGTGAATGGGTGTCT	382
QY	548	TGCGTTGA-----GAGTGGGTCCGGGCACGAAAGTTGGTGAGGGTTGAGAAAGCTGGAA	601
Db	383	TGTGTTGATAGGAGGAGTGGTTTGGGCATGAAACTTGGGTTAGAGCTGAGAAAGCTGGAG	442
QY	602	TTGGGGCTGTGTGGGAAGAGCAGGCCCGTTTCTTCAGAGGGTGGCGACTGTGGTGACG	661
Db	443	AAACCCCTTTGTGGGAAGAGTAGGCCCTGTTTTCTTTAGAGGGTTGCCACCATTTGTTACC	502
QY	662	AAGTTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTTAAGAGGATTATCAGCAG	721
Db	503	AAGTTGTTTAATATTGTGGAGCCTGATGTTGCTGTGTTGGGAAGAGGACTATCAGCAA	562
QY	722	TGGCGGCTTATTTCAGAGGATGGTTCGAGATCTTGATTTTTCCATAAAAGTGATAGGTGCT	781
Db	563	TGGAAAAATTATTTCAGAGAAATGGTTCGAGATCTTGATTTTTCCATAAAAGTGATAGGTCT	622
QY	782	GAAATAACACGTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTTCACCT	841
Db	623	GAAGTAATACGTGAGAAAGATGGCCTAGCAATGAGTTCGCCGTAATGTGTACCTATCACCT	682
QY	842	GAAGAGAGGGAAGGCACTATCAATAAATAATCATTTGTTAAGAGCTAAATCAGCAGCA	901
Db	683	GAAGAGAGGGAAGGCACTATCTATAATAATCATTTGTTTAGAGCTAAATCGGCAGCA	742
QY	902	GGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTGCATCCAAAGTTACTGAT	961
Db	743	GAAGATGGACAGATACATTGTGAGAAATTTGATAAACTTGGTCGTGCAAAAGTATCACCGAA	802
QY	962	GCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAAATTTGGAGAAAGTGGAA	1021
Db	803	GCTGGTGAAGGATTAATTATGCTGAGATTGTTGATCAAAAATAAATTTGGAGAAAGTGGAA	862
QY	1022	CAGATCAAGAGTCCGTGCTCTCTGTGTGCTGATGGTTGGCAAGTCAAGCTTATA	1081
Db	863	TGGATCAAGGGTCCGTGTCTTCTGTGTTTCTGATGGTTTGGGAAAGCCAGGCTTATA	922
QY	1082	GACAAACATGGAATCAACTTGTCAATGAATGT---TTGATCTAACCTTCTG--TCATCT	1135
Db	923	GACAACATAGRAATCAACTGTAAATGGAAGTAAGATTGATCTAACCTTGTGAATAATCT	982
QY	1136	CAAACATGGGCCACATGCTTAATTAAATAGTTCCGGGCCACGTGCTTAACAATTC	1188
Db	983	CAGACATGGACCATATGATTAGTAGTTCTTGGCATTTTCATGGGGTATAGACTTC	1035

RESULT 3				
A94671	LOCUS	A94671	1347 bp	DNA
	DEFINITION	Sequence 9 from Patent WO9942565.		
	ACCESSION	A94671		
	VERSION	A94671.1	GI:6778947	
	KEYWORDS	.		
	SOURCE	unidentified		
	ORGANISM	unidentified		
				linear
				PAT 26-JAN-2000

```

unclassified.
1 (bases 1 to 1347)
Laber,B. and Genschel,U.
HERBICIDES TEST METHOD
Patent: WO 9942565-A 9 26-AUG-1999;
LABER BERND (DE); GENSCHEL ULRICH (DE)
FEATURES
    Location/Qualifiers
        source          1..1347
                        /organism="unidentified"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32644"
        exon            1..1347
BASE COUNT      377 a    254 c    321 g    395 t
ORIGIN
Query Match      45.6%; Score 599.6; DB 6; Length 1347;
Best Local Similarity 76.8%; Pred. No. 1.4e-105;
Matches 805; Conservative 0; Mismatches 219; Indels 24; Gaps 5;

QY 252 CCCAGGGCAAGCTCATTTGGGCTGGTCCCAACCATGGGTTCTTCCACGCGGGCCACCTCT 311
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    69 CCCAAGGCAAGCTCATCGCCCTCGTTCCCAACCATGGGCTTCTTCCAGAGGCCACCTTT 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 312 CGCTCGTGGCCCGCCCGCCCAACTCTCCGACGTCGTCGCCGTCTCCATCTACGTCAACC 371
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    129 CTCCTCGTCAGAGACGCTCACAAACCACGCTGACCTCGTCGCCGTCTCAATCTATGTCAACC 188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 372 CGGGCCAATTCCGCCCCCAGGAGGACCTCTCCACCTACCCCTCCGACTTCGACGGCGACG 431
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    189 CTGGCCAGTTTCCCGCACCGAGGACCTTCCGCAATACCTTCTGATTTTCAAGGTGATC 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 432 TAAAGAAACTCGCGTCCGTTCCCGGGCGGTCGACGTCGTTTTTCCATCCCGTAACCTGT 491
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    249 TCCAAAAACTCATGTCTGTTCTCGTGGTGTGATGTTGTTTTCCACCCCCACAATTTGT 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 492 ACGATTACGGGAAGAACGGTGGTGACGTGGCAGAG-----GCTGGTGAATGTGT 545
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    309 ATGATTACGGTGGTGATGGCGGTGATGCTGTGGCGAGTGTGGTGTATGGGGTGTGT 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 546 CGTGCCTTGA-----GAGTGGTCCGGGCACGAAAGTTGGGTGAGGTTGAGAACTGG 599
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    369 CTTGTGTTGATAGGAGGAGTGGTTTGGGCATGAACCTGGGTTAGAGCTGAGAACTGG 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 600 AATTGGGGCTGTGTGGGAAGAGCAGGCCCGTTTCTTTCAGAGGGGTGGCGACTGTGTGA 659
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    429 AGAAACCCCTTTGTGGGAAGAGTAGGCCCTGTTTCTTTAGAGGGGTTCGCCACCATTTA 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 660 CGAAGTTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTTAAGAAAGGATTATCAGC 719
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    489 CCAAGTTGTTTAATATTGTGGAGCCTGATGTTGCTGTGTTTGGGAAGAAGGACTATCAGC 548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 720 AGTGGCGGCTTAATCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAGTGATAGGTG 779
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    549 AATGGAAATTAATCAGAGAATGGTTCGAGATCTTGATTTTCCATAAAAGTGATAGGT 608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 780 CTGAAATAACACGCTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTAC 839
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    609 CTGAAGTAATACGTGAGAAAGATGGCCTAGCAATGAGTTCGCGTAATGTGTACCTATCAC 668
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 840 CTGAAGAGAGGGAAGGCACATCAATAATAAATCATTTGTTAAGAGCTAAATCAGCAG 899
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    669 CTGAAGAGAGGGAAGGAGCAGTATCTATAATAAATAATCATTTGTTAAGAGCTAAATCAGCAG 728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 900 CAGGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTGTACTG 959
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    729 CAGAAGATGGACAGATACATTGTGAGAAATTGATAAACTTGGTCGTGCAAAAGTATCACCG 788
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 960 ATGCTGGTGAAGCATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGG 1019
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    789 AAGCTGGTGAAGGATTGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGG 848
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1020 AACAGATCAAGAGTCCTGTCTTCTGTGTTGCTGTCATGGTTTGGCAAGTCAGGCTTA 1079
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



Db	849	AATGATCAAGGGTCCTGTTGTTCTTCTGTGTTTCTGCAATGGTTGGGAAAGCCAGGCTTA	908
QY	1080	TAGACAACATGGAAATCAACTGTGCAATGAATGTTTGATCT--AACCTTCTGTCACTCA	1137
Db	909	TAGACAACATAGAAATCAACTGTAAATGGAAGTAAGATTTGTATCTTTAGTTGAACAT	968
QY	1138	AACATGGGCCACATGCTTTAAATTAATAGTTCGGGCCACGTGCTTAACAATTCTAACAGTTC	1197
Db	969	TGATCTAACCTTGTGAATAATCTCAGACATGGACCATATGATTAGTATTCTGGCAATTC	1028
QY	1198	ATGTTATAGTCATGACAATTTTTTTTCTGCCAGCCAT-ACATGATTACTTGTAGATGC	1256
Db	1029	ATGG-----GGTATAGACTTCATTTCTACAAGCCATGATATGACTACTTGTAGATGT	1079
QY	1257	ATTTTACCGCATATAAAATTCATGAG	1284
Db	1080	ATTTTACTACCTCATGAAATTCAGGAG	1107
RESULT 4			
AY085534			
LOCUS	AY085534	1188 bp mRNA linear	PLN 14-APR-2003
DEFINITION	Arabidopsis thaliana clone 156371 mRNA, complete sequence.		
ACCESSION	AY085534		
VERSION	AY085534.1	GI:21404244	
KEYWORDS	FLI CDNA.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 1188)		
AUTHORS	Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.		
TITLE	Full-length messenger RNA sequences greatly improve genome annotation		
JOURNAL	Genome Biol. 3 (6), RESEARCH0029 (2002)		
MEDLINE	22088475		
PUBMED	12093376		
REFERENCE	2 (bases 1 to 1188)		
AUTHORS	Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.		
TITLE	Full-length cDNA from Arabidopsis thaliana		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 1188)		
AUTHORS	Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA		
COMMENT	This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or LAer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.		
FEATURES	Location/Qualifiers		
source	1. .1188		
	/organism="Arabidopsis thaliana"		
	/mol_type="mRNA"		
	/db_xref="taxon:3702"		
	/clone="156371"		
	37. .969		
CDS			

		BASE COUNT	348 a	207 c	295 g	338 t	
		ORIGIN					
		Query Match	32.1%;	Score 422.2;	DB 8;	Length 1188;	
		Best Local Similarity	66.8%;	Pred. No. 2.5e-71;			
		Matches 623;	Conservative 0;	Mismatches 298;	Indels 12;	Gaps 1;	
QY	197	CCAAGGGTGATCTCCGACAAAGGCCTCGATCGGAGCTGGTCGCGCTCGATGGGGCCCCAG	256				
Db	43	CCAGAAGTAATCAGAGACAAAGACTCTATGAGAAAATGGTCTCGAGCCATGAGATCACAA	102				
QY	257	GGCAAGCTCATTTGGGCTGGTCCCCACCACCATGGGCTTCCTCCACGGGGCCACCTCTCGCTC	316				
Db	103	GGCAAGACCATAGGTTTAGTACCAACAATGGGTTACCTCCACGAAGGTCACTTATCTCTC	162				
QY	317	GTGGCCCCAGGCCCCCAACTCTCCGACGTCGTGCGCGTCTCCATCTACGTCAACCCGGGC	376				
Db	163	GTTCGTCAATCACTAGCCCTCACTGACGTACCGTCGTCTCAATCTACGTCAATCCAGGC	222				
QY	377	CAATTGCCCCCACGAGGACCTCTCCACCTACCCCTCCGACTTCGACGGGACGTAAAG	436				
Db	223	CAATTCTCTCCACTGAAGACCTCTCCACATACCCATCTGATTTCTCCGGCATCTCACT	282				
QY	437	AAACTCGCGTCCGTTCCGGCGGCGTCGACGTCGTTTTCATCCCCGTAACTGTACGAT	496				
Db	283	AAGCTCGCGGCTCTTTCCGGTGGTAAAGTCGTCGCTTTAAACCCAAAAACCTCTATGAT	342				
QY	497	TACGG-----GAAGAACGGTGGTGTGACGTGGCAGAGGCTGGTGAATGGTG	544				
Db	343	TACGGCGGTGAGACGAAGAAGATAAACGACGGTGGTGGTAATGGTGGGAGGTAGTGAGT	402				
QY	545	TCGTGCGTTGAGAGTCGGTCCGGGCACGAAAGTTGGGTGAGGGTTGAGAGAGCTGGAATTG	604				
Db	403	TGTGTGGAGGAAGGTGGTTAGGGCATGAGACTTGGATTAGGGTTGAGAGATTGGAGAAA	462				
QY	605	GGCGTGTGTGGGAACAGCAGGCCCCGTTTTTCTCAGAGGGGTGGCGACTGTGGTGACGAAG	664				
Db	463	GGTTTGTGTGGGAACAGTAGGCCCTGTGTTCTTTAGAGGTGTTGCTACTATTGTTACTAAG	522				
QY	665	TTGTTTAATATTGTGAGCCAGATGTGGCTGTGTTCCGTTAAGAAGGATTATCAGCAGTGG	724				
Db	523	CTTTTAAATATTGTTGAGCCTGATGTTGCTCTGTTGGTTAAGAAAGATTATCAACAATGG	582				
QY	725	CGGCTTATTCAGAGGATGGTTCGAGATCTTGATTTTTTCCATAAAAGTGATAGTGTGAA	784				
Db	583	AGGATTATACAGAGAATGGTTCGAGATCTTAATTTTGGTATTGAGATTGTAGGATCAGAT	642				
QY	785	ATAACACGTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTCACCTGAA	844				
Db	643	ATAGCTAGAGAAAAGATGGACTTGCATGATGAGTTCGAGAAATGTGCGCTCTTCAGATGAA	702				
QY	845	GAGAGGGAAAAGGCACCTATCAATAAATAATCATTTGTTAAGAGCTAAATCAGCAGCAGGA	904				
Db	703	GAAAGGCAAAGGGCTTGTCTATTAAGTAGGTCACTGGCCATGGCTTAAAGCTTCTGTTGCA	762				
QY	905	GATGGTCAGGTGCATTGTGAGAAGTTTGACAAATTTGGTCATCCAAAGTGTACTGATGCT	964				
Db	763	GAAGGGAAAACCAATTTGTGCAGAGCTTAAAGGATATGATCATTTCAACAAGTTGTTGGATCT	822				
QY	965	GGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAAATAATTTGGAGAAAAGTGGAACAG	1024				
Db	823	GCAGGACGAGTAGATTATGTCGAGATAGTTGACCAAGAAACTCTTTGAAGGAGTAGAAGAG	882				





/translation="MAAPREPEVIRDKAAMRAWRRRAEKGKTVAVVPTMGYLHQHGL  
SLISAAAAASADPVAIVVTIYVNPSPQFAPSEDLATYPSDFAGDLRLKLASTGWDAVF  
NPPDLIYVRGARRGAGSGGAI5CLEEAAAGDHETWVRVERLEKGLGSRPVPFFRGVA  
TIVSKLFNIIIEPDVPFGKKDYQQWRVILPYWGLDFGIEIMGSRNCARTDGLAMNSR  
NVHLSREEGKKALSISRSLVDARTGALKGNTDSKQIKNKIVQTLTETGGQVDYVEIVE  
QESLVPVEIQIDGPVVICVAAWFGKVRLLIDNIBIDTRS"

BASE COUNT		279 a	349 c	359 g	277 t
ORIGIN					
Query Match 26.9%; Score 354; DB 8; Length 1264;					
Best Local Similarity 63.2%; Pred. No. 3.7e-58;					
Matches 584; Conservative 0; Mismatches 325; Indels 15; Gaps 2;					
QY	187	GGTCCAGCCCCAAGGTTGATCTCCGACAAGGCCTCGATCGGAGCTGGTCGCGCTCGAT	246		
Db	83	GCCGCGAGCCGGAGGTGATCCGCGACAAGGCGGATGCGGCATGGTCGCGCCGCCG	142		
QY	247	GCGGGCCAGGGCAAGCTCATTCGGCTGGTCCCCACCATGGGCTTCCTCCACGCGGGCCCA	306		
Db	143	TGCGCCGAGGGCAAGACCGTCCGGTCTGTACCCACCATTGGGCTACCTCCACCAAGGCCA	202		
QY	307	CTCTCTCGCT-----CGTGGCCCCAGGCCCGCCCAACTCTCCGACGTCGTCGCGT	354		
Db	203	CTCTCCCTCATCTCCGCCGCCGCCGCCCTCCGCTCGCTGATCCCGTCGCCATCGCTCGT	262		
QY	355	CTCCATCTACGTCAACCCGGGCCAATTGCCCCCACCAGGAGACCTCTCCACCTACCCCTC	414		
Db	263	CACCATCTACGTCAACCCAGCCAGTTTCGCGCCCTCAGAGGACCTCGCCACCTACCTTC	322		
QY	415	CGACTTCGACGGCGACGTAAAGAAACTCGCGTCCGTTCCCGCGGCGTTCGACGTCGTTT	474		
Db	323	CGACTTCGCCGGTGACCTCCGCAAGCTCGCCTCC--ACCGGCGTCGTGGATGCCGCTT	379		
QY	475	CCATCCCCGTAACTTGTACGATTACGGGAAGAACGGTGGTGGTGACGTGGCAGAGGGTGG	534		
Db	380	CAACCCCTTGACCTCTACGTCCGTGGCGCCGGTCCGCCGGGCGCGGTCCGGAGCGC	439		
QY	535	TGGAATGGTGTGTCGTTGAGAGTGGTCCGGGCACGAAAGTTGGTGAGGGTTGAGAA	594		
Db	440	GATCTCTGCTCGTAGGAGGCGCGCGGGATGGGCACGAGACGTGGGTTCCGGTGGAGCG	499		
QY	595	GCTGAATTGGGCTGTGTGGGAAGACAGGCCCGCTTTCTTCAGAGGGTGGCGACTGT	654		
Db	500	ATTGGAGAAGGGATTGTGCGGGGCCAGCCGTCCTCGTGTCTTCCTCCGAGCGTGGCCACAT	559		
QY	655	GGTGACGAAGTTGTTAATATTGTGAGCCAGATGTGGCTGTGTTCGGTAAGAAGGATTA	714		
Db	560	AGTCTCAAGCTGTTTAAACATCATCGAGCCGGATGTTCTGTGTTCGGGAAGAAGGATTA	619		
QY	715	TCAGCACTGGCGCTTATTTCAGAGGATGGTTCGAGATCTTCATTTTCCATAAAATGAT	774		
Db	620	TCAGCAGTGGCGCTCATCTTGCCGTATTGGTCGGGACITGATTTTGGCATAGAGATAAT	679		
QY	775	AGTGTCTGAAATAACACGTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCT	834		
Db	680	GGGATCAAGAAATTGTGCGAGAACTGATGGTCTTGCCATGAACCTCCCGAATGTGCACCT	739		
QY	835	TTCACTGAAGAGAGGGGAAAAGGCACTATCAATAATAATCATTTGTTAAGAGCTAAATC	894		
Db	740	ATCACCGAGGAAGGGAAGGCAATATCCATCAGTAGATCACTGGTTGATGCTAGAAC	799		
QY	895	AGCAGCAGGAGATGGTCAGGTGCAATTGTGAGAAGTTGACAAATTTGGTCATCCAAAGTGT	954		
Db	800	TGGCGCCTGAAGGGAAACACTGATTCCAAACAATCAAAAACAAAATAGTACAGACACT	859		
QY	955	TACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAA	1014		
Db	860	AACTGAAGTGGCGGTCAGGTTGACTATGTTGAGATCGTGGAGCAAGAAAGTTTGGTCCC	919		
QY	1015	AGTGGACAGATCAAGAGTCCTGTCGTCTTCTGTGTTGTCGATGTTTGGCAAAGTCAG	1074		
Db	920	TGTAGAACAGATCGACGGCCCTGTGCTCATTTGCGTTTGGCGCTGGTTTGGAAAGGTCAG	979		

QY	1075	GCTTATAGACACATGGAAATCAA	1098		
Db	980	GCTGATCGATATATCGAAATCGA	1003		
RESULT 7					
AB012242/c					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
Arabidopsis thaliana (thale cress)					
Arabidopsis thaliana					
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
1 (sites)					
Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Miyajima,N.					
and Tabata,S.					
Structural analysis of Arabidopsis thaliana chromosome 5. VI.					
Sequence features of the regions of 1,367,185 bp covered by 19					
physically assigned P1 and TAC clones					
DNA Res. 5 (3), 203-216 (1998)					
98403884					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
Direct Submission					
Submitted (23-MAR-1998) Yasukazu Nakamura, Kazusa DNA Research					
Institute, Department of Plant Gene Research; 1532-3, Yana,					
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,					
Tel:81-438-52-3935, Fax:81-438-52-3934)					
Address for correspondence: kaos@kazusa.or.jp					
For the latest information on annotation of this clone, please see					
http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=k24G6					
Genes with similarity to proteins in the databases are described in					
'product' or 'note' qualifiers. Genes that have no significant					
protein similarity are described as 'unknown protein'.					
The software programs used to predict genes include: Grail					
(Informatics Group, Oak Ridge National Laboratory,					
http://compbio.ornl.gov/Grail-1.3/),					
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),					
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of					
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and					
SplicePredictor (Volker Brendel, Stanford University,					
http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi).					
Genes encoding tRNAs are predicted by tRNAscan-SE					
(Sean Eddy, Washington University School of Medicine, St. Louis,					
http://genome.wustl.edu/eddy/tRNAscan-SE/).					
This sequence may not be the entire insert of this clone. It may be					
shorter because we remove overlaps between neighboring submissions.					
The 5' clone is K15N18 and the 3' clone is K19E20.					
FEATURES					
source					
1..78973					
/organism="Arabidopsis thaliana"					
/mol_type="genomic DNA"					
/strain="Columbia"					
/db_xref="taxon:3702"					
/chromosome="5"					
/clone="K24G6"					
/clone_lib="Mitsui TAC"					
complement(1..49)					
/note="CDS is reported in Acc# AB015468					
gb AAF20229.1					
gene_id:K15N18.18					
similar to unknown protein"					
/number=5					
/evidence=not_experimental					
complement(144..215)					
/note="CDS is reported in Acc# AB015468					
gb AAF20229.1					
gene_id:K15N18.18					
exon					
exon					







DEFINITION Oryza sativa chromosome 3 BAC OSJNBa0015N08 genomic sequence, complete sequence.

ACCESSION AC096688

VERSION AC096688.4 GI:27228823

KEYWORDS HTG.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 145290)

AUTHORS Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M., Overton II,L.L., Tsitrin,T., Kim,M.M., Bera,J.J., Jin,S.S., Fadrosh,D.W., Tallon,L.J., Koo,H., Zismann,V., Hsiao,J., Blunt,S., Vanaken,S.S., Riedmuller,S.B., Utterback,T.T., Feldblyum,T.V., Yang,Q.Q., Haas,B.J., Suh,B.B., Peterson,J.J., Quackenbush,J., White,O., Salzberg,S.L. and Fraser,C.M.

TITLE Oryza sativa chromosome 3 BAC OSJNBa0015N08 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 145290)

AUTHORS Buell,R.

TITLE Direct Submission

JOURNAL Submitted (22-SEP-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 145290)

AUTHORS Buell,R.

TITLE Direct Submission

JOURNAL Submitted (18-DEC-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 4 (bases 1 to 145290)

AUTHORS Buell,R.

TITLE Direct Submission

JOURNAL Submitted (10-JAN-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuella@igr.org

COMMENT On Dec 18, 2002 this sequence version replaced gi:18032992. Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0015N08 is from Oryza sativa chromosome 3

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh (<http://www.softberry.com/>), Genscan and Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

This BAC overlaps with rice BAC OSJNEb0096M04 (AC092559) and OSJNEb006O08 (AC120506).

Location/Qualifiers

1. .145290

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="3"

/map="near L375"

/clone="OSJNBa0015N08"

/note="japonica cultivar-group"

complement(116. .1322)

/gene="OSJNBa0015N08.1"

/note="EST C72581 from this gene"

complement(join(<116. .351,931. .1130,1204. .>1322))

/gene="OSJNBa0015N08.1"

complement(join(<116. .351,931. .1130,1204. .1322))

/gene="OSJNBa0015N08.1"

/codon\_start=1

/product="unknown protein, 3'-partial"

/protein\_id="AAO20052.1"

/db\_xref="GI:27573334"

/translation="MLNESMGEGDAAYAKRVLLTAAGDDVSRGIASLTATHGCRLLVLV GDEGLAGTAEERARRGGGGDAVAVVGLDLHGCDEAAVDAVGTAWRCFDGLDAMVNC VSYEGEVQDCLNISEDEFKTKMKANVMTWFLVKAIAKRLRDSSESSCGGSVVFLTQII GAERGLYPGAAAYGTSLGAIHQLVR"

1399. .1431

/rpt\_family="AT\_rich"

1453. .1515

/rpt\_family="AT\_rich"

1594. .1617

/rpt\_family=" (TTTTc)n"

complement(2130. .5799)

/gene="OSJNBa0015N08.2"

/note="contains Pfam profile PF03031 (NLI interacting factor)"

complement(join(<2130. .2324,3346. .3683,4248. .4676, 4905. .5001,5205. .5329,5649. .>5799))

/gene="OSJNBa0015N08.2"

complement(join(2130. .2324,3346. .3683,4248. .4676, 4905. .5001,5205. .5329,5649. .5799))

/gene="OSJNBa0015N08.2"

/codon\_start=1

/product="putative NLI interacting factor"

/protein\_id="AAO20054.1"

/db\_xref="GI:27573336"

/translation="MPTRKKGUARNATTEHINPNSRKSRSRSTQAAAAEHKVNDLITS SSKKQIGGGLLKQRALRGKKLNSIYDSAGTNDVAGVPSSSFFNHHKQSHENDEDRS CDSIFSPSFHNQKEDVTDCLSEGLVHEEVIHEPQVNAEYVNSISCNTFDGLSKHSYD IHMQSACGSTLEGDEFSELGSLSPESVAIYLAMQHSKLECIDESQSDSTSTDGCADPD ETVELDYDFPYAFIKDPLSLVVPKQPVLPLPKQTRCPTTLVLDDLDETLLVHSTLE PCEDSDFTFPVHFNLRHTIYVRCRPLYKEFLETVASMEFIIIFTASQSIYAEQLLNI LDKRRFLRHRVYRESCLFVEGNLYKDLVLGRDLARVVVDNSPQAFGFQLDNGVPI ESWFDDRNDRELLTLLPFLQSLVGVEDVRPCIARKFNLRKVATAPSLSMHF"

6846. .6880

/rpt\_family="AT\_rich"

7034. .10067

/gene="OSJNBa0015N08.3"

/note="similar to cytochrome P450 CYP71E1 GB:AAC39318 GI:2766452 (Sorghum bicolor)"

join(<7034. .7534,8885. .9343,9444. .>10067)

/gene="OSJNBa0015N08.3"

join(7034. .7534,8885. .9343,9444. .10067)

/gene="OSJNBa0015N08.3"

/codon\_start=1

/product="putative cytochrome P450 protein"

/protein\_id="AAO20056.1"

/db\_xref="GI:27573338"

/translation="MATLSKLLALPQQWQLLLLLLLPIASLLLVIGRNTGRRRRR HLRLPPGPARLPVLGNLLQLGALPHRSLRDLARRHGPVWMLRLGAVPAVVVSSPEAAQ EVLRTHDADCCSRPSSPGMRLSYGYKDAFAPYDAYGRAARRLFVAELFSAPRVQAA WRARQDQVEKLIKGLTRPEPEPVELNDHIFALTDGIIGAVAFGSIYGTFRFAGGGRKR FHLLDDVMDMLASFSAEDFFPNAARLFDHLTGLVAHRERVFQLDAFFEMVIEQH LWDSSNAGGGGNLVGALIGLWKQGYGDRRFRFTRNNVKAIIFDAFIGIGTSSVTI LWAMAEMLRSPRVMRKVQAEIRATVGDRDGGWVQDDLPRLAYLKMVVKETLRLHPF ATLLMPRETRMDVRIGGYEVAARTVMVNAWAIGRDAARWEEAEVDPDRFEAKRVFF NGCHFELLPFSGGRRICPGIANGAANVEFTLANLLHCFDWALPVGMAPEELSMERSGG LVLHRKAPLVLPTRYQL"

7082. .7106

/rpt\_family=" (TCC)n"

8481. .8595

/rpt\_family="AT\_rich"

10171. .10217

/rpt\_family="AT\_rich"

complement(14707. .15477)

gene

repeat\_region

repeat\_region

repeat\_region

gene





```

MIRMPVSEIGVIGRVTKGVKLIELGDDTTISKVAVVKD"
792..1385
/gene="lexA"
792..1385
/gene="lexA"
/EC_number="3.4.21.88"
/codon_start=1
/transl_table=11
/product="LexA repressor"
/protein_id="CAA07513.1"
/db_xref="GI:3282090"
/db_xref="SPTREMBL:O86948"
/translation="MKDLTAKQRSVLITFEEFIEKNKY
FLVALEKGYIERKNGKPRAMRVTKSVKNRVPILGEIR
HLSSGYEHPFLRVKKGESMIEEHI CDGDLVLRQDWAQ
RGEMVELEPANKEMSPMFFRADRVKILGVKVGVFRKI"

```

BASE COUNT		1309 a	845 c	1228 g	901 t
ORIGIN					
<pre> /gene="ORF7" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAA07518.1" /db_xref="GI:3282095" /db_xref="SPTREMBL:O86953" /translation="MRIIETIEMKKFSEEMREKKKTIGFVPTMGYLHEGHLSLVRRRA RAENDVVVVSIFVNPTQFPNEDYERYPRDFERDKLLEKENYDVCVFHPSVEEMYPPD FSTVEETKLSKPLCGRSRPHGFGVCTVTKLFNIKPHRAYFGQKDAQFRVLRRM VRDLNMDVEMIECPIVREPDGLAMSSRNVYLTPEERKQALALYQSLKIAENLFLNGER DAVKIKEAMINHLRSFRDVKIDYVEIVDEETLEPVEKIDRKVIIVAVAAVWGKARLIDN TILG" </pre>					
Query Match		11.2%;	Score 147.6;	DB 1;	Length 4283;
Best Local Similarity		57.1%;	Pred. No. 3e-18;		
Matches 289;		Conservative 0;	Mismatches 214;	Indels 3;	Gaps 1;
QY	593	AAGCTGGAATTGGGGCTGTGTGGGAAGAGCAGGCCCGCTTTTCTTCAGAGGGGTGGCGACT	652		
Db	3760	AAGCTTTCAAAACCGCTGTGTGGTAGATCCAGGCCCGGACACTTTTCAGGGGTGTGTACC	3819		
QY	653	GTGGTGACGGAAGTTGTTTAAATATTGTGGAGCCAGATGTGGCTGTTCGGTAAGAAGGAT	712		
Db	3820	GTGGTCACAAGACTCTTCAACATCGTGAAACCAACACAGGGCGTACTTCGGTCAAAAAGAT	3879		
QY	713	TATCAGCAGTGGCGGCTTATTTCAGAGATGGTTCGAGATCTTGATTTTCCATAAAAGTG	772		
Db	3880	GCACAGCAGTTCAGGGTACTCAGGAGAATGGTGAGAGATCTGAACATGGACGTGAGATG	3939		
QY	773	ATAGGTGCTGAAATAAACACGTGATATGATGGCCTGGCAATGAGTTACGTAATGTGCAC	832		
Db	3940	ATCGAGTGTCCTCATAGTGAGAGAACCCAGACGGTCTTGGATGAGTTCAGAAACGTCTAT	3999		
QY	833	CTTTACCTGAAGAGAGAGGGAAAAGGCACATATCAATAATAATCAATTGTTAAGAGCTAAA	892		
Db	4000	CTCACACCTGAAGAGAGAAAACAGGCTCTTGCCCTGTACCAGTCGTTGAAAAATCGTGAA	4059		
QY	893	TCAGCAGCAGGAGATGGTCAGGTGCATGTGAGAAAGTTGACAAATTTGGTCATCCAAAGT	952		
Db	4060	AACCTCTTTTGAACGGTGAGAGGGATGCTGTGAAGATAAAGGAACGATGATCAATCAT	4119		
QY	953	GT---TACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTG	1009		
Db	4120	CTGTCCAGATTCGACAGAGTGAAGATAGACTACGTGGAGATCGTCGATGAGAAACGCTG	4179		
QY	1010	GAGAAAGTGAACAGATCAAGAGTCTCTGTGCTCTTCTGTGTTGCTGCATGTTTGGCAA	1069		
Db	4180	GAGCCCGTTGAAAAGATCGATCGAAAGGTGATCGTCGCGGTTCGCCCTGGGTGGGAAAG	4239		
QY	1070	GTCAGGCTTATAGACAACATGGAAAT	1095		
Db	4240	GCAAGACTGATAGACAAACACGATACT	4265		

RESULT	11
AE001768/c	
LOCUS	
DEFINITION	19711 bp DNA linear BCT 02-JUN-1999
ACCESSION	Thermotoga maritima section 80 of 136 of the complete genome.
VERSION	AE001768 AE000512
KEYWORDS	AE001768.1 GI:4981619
SOURCE	.
ORGANISM	Thermotoga maritima
	Thermotoga maritima
	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
REFERENCE	1 (bases 1 to 19711)
AUTHORS	Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,

Eisen,J.A., Fraser,C.M. et al.  
Evidence for lateral gene transfer between Archaea and bacteria  
from genome sequence of Thermotoga maritima  
Nature 399 (6734), 323-329 (1999)  
99287316  
PUBMED  
10360571  
REFERENCE  
2 (bases 1 to 19711)  
Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,  
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,  
McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,  
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,  
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,  
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.  
Direct Submission  
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
1..19711  
/organism="Thermotoga maritima"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:2336"  
complement(67..978)  
/gene="TM1075"  
complement(67..978)  
/gene="TM1075"  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAD36152.1"  
/db\_xref="GI:4981620"  
/translation="MKDGKLYVAFSGGKDSLSVAIIAKMALGEERVELVTVDWSPYTY  
ERSREIVRNFAEKHGLKHTFIPSNRMQEKVMKHGSPSCNACTRDVKTIVLKRYAQGHLV  
ASGANASDSWGKTGLKVPFDGVYSPLCRVGKEINEMLKFLGLEVKKIGESAGREGCKL  
KHLKMLINPDYHGKAVSTANEILLRVLEEHGFKPELANVKIIGPLSRNIALVNVKPL  
PPEKVMNEIIVEKLSAEETIDGVIIVDGPMLVVLASPAIYRNEESRXWIKEGRLQPEF  
AFPIEIEWRESKNNKLRTFQVVDARKE"  
complement(1008..1481)  
/gene="TM1076"  
complement(1008..1481)  
/gene="TM1076"  
/note="similar to GB:AE000782 percent identity: 51.61;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36153.1"  
/db\_xref="GI:4981621"  
/translation="MKRFLIISDSHVPRMASLPDEILNSLKEYDGVIGLDYVDLDT  
VILKEKFSKEFYGVHGNDYDVKHLPFSKVLLVEGVTIGMCHGWGAPWDLKDRLLK  
VFNEKPQVILFGHTHEPEDTVKAGVRFLNPGSLAEGSYAVLELDGGEVRFELKTL"  
complement(1486..2328)  
/gene="TM1077"  
complement(1486..2328)  
/gene="TM1077"  
/note="similar to GB:AE000657 percent identity: 74.18;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="pantoate--beta-alanine ligase"  
/protein\_id="AAD36154.1"  
/db\_xref="GI:4981622"  
/translation="MRIETIEEMKKFSEEMREKKKTIGFVPTMGYLHEGHLSLVRRRA  
RAENDVVVSIFVNPQTQFGPNEDYERYPRDPFRDKLLEKENVDCIFPHPSVEEMYPPD  
FSTYVEETKLSKILCGRSRPGHFRGVCVTVKLFNIVKPHRAYFGQKDAQQFRVLRRM  
VRDLNMDVEMIECPIVREPDLGAMSSRNVYLSPEERQQALSLYQSLKIAENLYLNGER  
DAEKIKEIMIKHLSRFDKVKIDYVEIVDEETLEPVEKIDRKVIVAAWVGNARLIDN  
TILG"  
complement(2325..2837)  
/gene="TM1078"  
complement(2325..2837)  
/gene="TM1078"

/note="similar to GB:AE000657 percent identity: 65.74;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36155.1"  
/db\_xref="GI:4981623"  
/translation="MGDPREKIVGQIGRKRKITNMRRVVFRCPYGVVPIESFPIKDG  
KPFPTLYMTCPHLRKEVSKLESQGVKEFEKLEREGKLRKMKKAHVEIIKKRAEL  
LPEDHPPREILSKVGTGGIRDFTKLKCLHLHLADYLAGVKNPVGEMVMDLVERKFCEE  
KLCRAGEARE"  
complement(2807..3601)  
/gene="TM1079"  
complement(2807..3601)  
/gene="TM1079"  
/note="similar to GB:AE000657 percent identity: 61.19;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal protein L11 methyltransferase,  
putative"  
/protein\_id="AAD36156.1"  
/db\_xref="GI:4981624"  
/translation="MRFKELILPLKIEEELVEKFYEEGFNFAIEEDKKGKVLKIY  
LREGELPDLFKDWEIVDEKITTPKDWIWELEPFEIVEGIFIDPTEKINRRDAIVIKL  
SPGVAFGTGLHPTTRMSVFFLKYYLKEGNTVLDVCGGTGILAAAKKLGASRVVAVDV  
DEQAVEVAEENVRKNDVDVLVKSDLLSEVEGTFDIVVSNILAEIHVKLLEDVNRVTH  
RDSMLILSGIVDRKEDMVKRKASBEHGMNVLERKQEREWTLVMKRS"  
complement(3591..4022)  
/gene="TM1080"  
complement(3591..4022)  
/gene="TM1080"  
/note="similar to GB:AE000657 percent identity: 77.62;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="sugar-phosphate isomerase"  
/protein\_id="AAD36157.1"  
/db\_xref="GI:4981625"  
/translation="MKIAIASDHAAFELKEKVKNYLLKGIEVEDHGTSEESVDYPD  
YAKKVQSILSNEADFGILCGTGLGMSIAANRYRGIRAAALCLFPDMARLARSHNAN  
ILVLPGRLLGAELAFWIVDTFLSTPFDGGRHERRIRKIDEV"  
complement(4019..4360)  
/gene="TM1081"  
complement(4019..4360)  
/gene="TM1081"  
/note="similar to GP:2660753 percent identity: 64.15;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="anti-sigma factor antagonist, putative"  
/protein\_id="AAD36158.1"  
/db\_xref="GI:4981626"  
/translation="MFPYKIVDDVWVILMPNKNELNIENAHLPKKWVFDEFLNKGYNKIF  
LVLSDVESIDSFSLGVI VNILKSISSSGGFFALVSPNEKVERVLSLTNLDRIVKIYDT  
ISEAMEEVRRK"  
complement(4372..4965)  
/gene="TM1082"  
complement(4372..4965)  
/gene="TM1082"  
/note="similar to GP:2645227 percent identity: 100.00;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="lexA repressor"  
/protein\_id="AAD36159.1"  
/db\_xref="GI:4981627"  
/translation="MKDLTERQRKVLFFIEEFIEKNGVPPSVREIARRFRITPRCALL  
HLIALEKKGYIERKNGKPRALRISKIRNKIPLIGETRAGEKREAIEYLEDYIEIPES  
FLSSGYDHFLLKVKGESMIEEHCDDGLVLVRRQDWAQNGDIVAAMVDGEVTLKKFYQ  
RGDTVELRPPANREMSSMFFRAEKVKILGKVVGVFRKL"  
complement(4962..5336)  
/gene="TM1083"

Eisen,J.A., Fraser,C.M. et al.  
Evidence for lateral gene transfer between Archaea and bacteria  
from genome sequence of Thermotoga maritima  
Nature 399 (6734), 323-329 (1999)  
99287316  
PUBMED  
10360571  
REFERENCE  
2 (bases 1 to 19711)  
Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,  
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,  
McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,  
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,  
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,  
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.  
Direct Submission  
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
1..19711  
/organism="Thermotoga maritima"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:2336"  
complement(67..978)  
/gene="TM1075"  
complement(67..978)  
/gene="TM1075"  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAD36152.1"  
/db\_xref="GI:4981620"  
/translation="MKDGKLYVAFSGGKDSLSVAIIAKMALGEERVELVTVDWSPYTY  
ERSREIVRNFAEKHGLKHTFIPSNRMQEKVMKHGSPSCNACTRDVKTIVLKRYAQGHLV  
ASGANASDSWGKTGLKVPFDGVYSPLCRVGKEINEMLKFLGLEVKKIGESAGREGCKL  
KHLKMLINPDYHGKAVSTANEILLRVLEEHGFKPELANVKIIGPLSRNIALVNVKPL  
PPEKVMNEIIVEKLSAEETIDGVIIVDGPMLVVLASPAIYRNEESRXWIKEGRLQPEF  
AFPIEIEWRESKNNKLRTFQVVDARKE"  
complement(1008..1481)  
/gene="TM1076"  
complement(1008..1481)  
/gene="TM1076"  
/note="similar to GB:AE000782 percent identity: 51.61;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36153.1"  
/db\_xref="GI:4981621"  
/translation="MKRFLIISDSHVPRMASLPDEILNSLKEYDGVIGLDYVDLDT  
VILKEKFSKEFYGVHGNDYDVKHLPFSKVLLVEGVTIGMCHGWGAPWDLKDRLLK  
VFNEKPQVILFGHTHEPEDTVKAGVRFLNPGSLAEGSYAVLELDGGEVRFELKTL"  
complement(1486..2328)  
/gene="TM1077"  
complement(1486..2328)  
/gene="TM1077"  
/note="similar to GB:AE000657 percent identity: 74.18;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="pantoate--beta-alanine ligase"  
/protein\_id="AAD36154.1"  
/db\_xref="GI:4981622"  
/translation="MRIETIEEMKKFSEEMREKKKTIGFVPTMGYLHEGHLSLVRRRA  
RAENDVVVSIFVNPQTQFGPNEDYERYPRDPFRDKLLEKENVDCIFPHPSVEEMYPPD  
FSTYVEETKLSKILCGRSRPGHFRGVCVTVKLFNIVKPHRAYFGQKDAQQFRVLRRM  
VRDLNMDVEMIECPIVREPDLGAMSSRNVYLSPEERQQALSLYQSLKIAENLYLNGER  
DAEKIKEIMIKHLSRFDKVKIDYVEIVDEETLEPVEKIDRKVIVAAWVGNARLIDN  
TILG"  
complement(2325..2837)  
/gene="TM1078"  
complement(2325..2837)  
/gene="TM1078"

CDS  
complement(4962. .5336)  
/gene="TM1083"  
/note="similar to GB:L77117 SP:Q60334 PID:1498783 percent identity: 53.91; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36160.1"  
/db\_xref="GI:4981628"  
/translation="MRKPIEHTADIAYEISGNSYEBELLEARNILLEEEGIVLDTTEK EKMYPLEETEDAFDFTVNDWILEISKGWAPWRIKREGNELKVTFRKIRKEGTEIKAL TYHLLKXFERDGVLTQKVVFDT"  
complement(5333. .7747)  
/gene="TM1084"  
complement(5333. .7747)  
/gene="TM1084"  
/note="similar to GP:2645226 percent identity: 99.88; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase, subunit A"  
/protein\_id="AAD36161.1"  
/db\_xref="GI:4981629"  
/translation="MPEILINKPVEDELVESVLLYSMSVIVGRAIPDVRDGLKPVQRR ILYGMYELGLKHNPTKKSARIVGEVMKYHPHGDAPVYDALVRMAQPFMTMYPLIEG QGNFGSIDRDPAPAMRYTEARLRLAEEMLEIDIEKNTVMINDNFDGLTKEPEVLPSKV PNIINGASGIAVGMATNIPPHNLSETDALIYLIDHPEATEVELMQFIKGPDFPTGA VVUNASELKKVVEEGRGRIIVRGKVHVEDGKRVKRIVITEIPYGVSKAGLIEQIAKIA KDDSELPIRNIRDESDKRGMRIVIEIPKDANEVEIINLYKRTALQDYFNVQMLVIDK HKRPLRLNLKGLMEAFLEHRFEVIRRRARYEYEQYTRRAHVVEGLLKAARAIGVVVDI VRNSKVESARQSLMETLEITEEQAKALIDMRLSRLTSLEIENLQNEYSDLVRKISEV KEILKDEKVEIMKKEFLYLKQYGDPRRTEVTDQSIENEELIVEEDVVITLSHK GYLKSTPLNSYRSQKRGKGITVSKLSEDEVEFVVAKNTSSTLFTINLGRAYVLKN YQLETTGRNTRGRHITAFNLNDETEKIVALASLNGEGRDLVIATKSGIKRKTALKEFE NATSNRGVRAIKIEPGDEIVSARVNSEKETLIVATKMGMAIRFVSDVRMRGNAAG VQAIKLQGDVEVSDVIPPGEEGEILTVEKFGKRTFPQLYRIQRRGGTGLRNISD VNKTGYVVAVRYVRGDEEIVVVTRNGMMIRFPVSEIGVIGRVTKGVKLIELGDDTISK VAVVKD"  
complement(7761. .9650)  
gene

Query Match 10.3%; Score 135.6; DB 1; Length 19711;  
Best local Similarity 55.8%; Pred. No. 7.3e-16;  
Matches 280; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

QY 592 GAAGCTGGAATTGGGCTGTGTGGGAAGACGAGCCCGTTTCTTCAGAGGGGTGGCGAC 651  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1999 GAAACTTTCAAAACACCTGTGTGGTAGATCCCGTCTGGTCAATTCCTGGAGTGTGCAC 1940  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 652 TGTGGTGACGAAGTTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGA 711  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1939 CGTGGTCAAGAGCTCTTCAACATCGTGAACCTCACAGGGCGTACTTCGGTCAAAAAGA 1880  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 712 TTATCAGCAGTGGCGCTTATTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAAGT 771  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1879 TGCACAGCAGTTCAGGGTTCTGAGGAGGATGGTGAGAGATCTGAACATGGACGTGGAAT 1820  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 772 GATAGGTGCTGAAATAACACGTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCA 831  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1819 GATCGAGTGTCCCATCGTGAGAGAACCCGATGGCCTCGCGATGAGCTCAAGAAACGTTTA 1760  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 832 CCTTTCACCTGAAGAGAGGGGAAAAGGCACTATCAATAATAATCAATGTTAAAGAGCTAA 891  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1759 TCTGTGCCCCGAAGAAAGGCAGCAGGCGCTTTCGCTCTATCAGTCTCTGAAAAATAGCGGA 1700  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 892 ATCAGCAGCAGGAGATGGTCAGGTGCATTCGTGAGAAAGTTGACAAATTTGGTCATCCAAG 951  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1699 GAATCTCTATCTGAACGGTGAAGAGACGCGAGAAAAGATAAAGAGAGATGATAAAACA 1640  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 952 TGGT---ACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTT 1008  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1639 CCTTTCCAGGTTTCGATAAGGTGAAGATAGACTACGTGGAGATCGTGGACGAGGAGACTCT 1580  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1009 GGAGAAAGTGGACACATCAAGATCCTGTCTGTCTTCTGTGTTGCTGCATGTTGGCAA 1068  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1579 GGAGCCGGTTGAGAGATCGATCGAAAGGTGATCGTCGCGTGGCTGCCGTGGGTTGAAA 1520  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 1069 AGTCAGGCTTATACACACATG 1090  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1519 TGCAAGACTCATAGATAACACG 1498  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
RESULT 12  
BX321856 302050 bp DNA linear BCT 23-APR-2003  
LOCUS Nitrosomonas europaea ATCC 19718, complete genome; segment 1/10.  
DEFINITION BX321856 AL954747  
ACCESSION BX321856.1 GI:30179916  
VERSION complete genome.  
KEYWORDS Nitrosomonas europaea ATCC 19718  
SOURCE Nitrosomonas europaea ATCC 19718  
ORGANISM Nitrosomonas europaea ATCC 19718  
Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales; Nitrosomonadaceae; Nitrosomonas.  
REFERENCE 1  
AUTHORS Chain,P., Lamerdin,J., Larimer,F., Regala,W., Land,M., Hauser,L., Hooper,A., Klotz,M., Norton,J., Sayavedra-Soto,L., Arciero,D., Hommes,N., Whittaker,M. and Arp,D.  
TITLE Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph Nitrosomonas europaea  
JOURNAL J. Bacteriol. 185 (9), 2759-2773 (2003)  
PUBMED 12700255  
REFERENCE 2 (bases 1 to 302050)  
AUTHORS Larimer,F.  
TITLE Direct Submission  
JOURNAL Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas genome consortium, the DOE Joint Genome Institute, Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, and the Genome Analysis Group, Oak Ridge National Laboratory, 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA; larimerfw@ornl.gov  
REMARK Nitrosomonas genome consortium  
FEATURES  
Location/Qualifiers  
1. .302050  
/organism="Nitrosomonas europaea ATCC 19718"  
/mol\_type="genomic DNA"  
/strain="ATCC 19718"  
/db\_xref="taxon:228410"  
211. .1590  
/gene="dnaA"  
/locus\_tag="NE0001"  
211. .1590  
/gene="dnaA"  
/locus\_tag="NE0001"  
/function="InterPro IPR001687:IPR001957:IPR003593 COG593"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="dnaA; chromosomal replication initiator protein"  
/protein\_id="CAD83912.1"  
/db\_xref="GI:30179917"  
/translation="MOKIETFWHFCLKHFRQELNGQFNWIKPLKLEVCPEKNTLI LIAPNRFVLQWIKDNFVTRIDEMAQDHFNERISFLELREPAESEAOQVRTSAQKNRE DKKPAAEKTQGTLSRKNPSQLNASFTDFAVTGKANQLARAGAIQVAERPGLAYNPL FIYGVGLGKTHLMQAINGVNLELDAGAKIRYVHAEKVSDVVSAYQHKSFDPKFLYY HSLDLLLVDDVQFFSGKNRTQEEFFYAFNALIEAHKQVIITSDCYPKEISGLEERLVS RFGWGLTVAIEPPELEMRVAILLKKALAEKJELDENTAFFIAKYIRSNVRELEGALKR VLAFSRFTGHSISLDLAKEALKDLLAIQNRQISINIENIQKTVDYKYIKVADMYSKRV RTIVRPQVAMATAKELTQLSLPDIGEAFGGRDHTTVLHAHRKIIELRTSDPGINRDF NALMHILRG"  
1791. .2912  
/gene="dnaN"  
/locus\_tag="NE0002"  
1791. .2912  
/gene="dnaN"  
/locus\_tag="NE0002"

gene  
CDS















SLVICVINIGQLPHVIWSIFESAFGWQEAAGAAGYTLQAITNGFORSFMSNEAGMG  
STPNAAAAAASWPHPAQAQIVQMIGIFIDTLVICITASMLILLAGNTYMPLEGIQ  
LIQKAMRVLMGSAEFVTLVILFAFSSIVANYIYAENNLFFLRNNPKAIWCLRIC  
TFATVIGGTLTSLPLMWQLADIIACMAITNLTAIILLSPVVHTIASDYLQRKLGVR  
PVFDPLRYPDIGRQLSRDAWDDVSQE"

Query Match		9.0%;	Score 118.8;	DB 1;	Length 281530;
Best Local Simlarity		48.3%;	Pred. No. 1.6e-12;		
Matches 427;		Conservative 0;	Mismatches 397;	Indels 60;	Gaps 1;
QY	239	CGCTCGATCGGGCCCCAGGGCAAGCTCATTTGGGCTGGTCCCCACCATGGGCTTCTCTCCAC	298		
Db	153344	CGCCGCCTGCGTATGGAAGGCAAGCGGTGGCGCTGGTACCTACCATGGGCAACCTGCAT	153285		
QY	299	GCGGGCCACCTCTCGCTCGTGGCCCCAGGCCCGCCAACTCTCCGACGTCGTCGCCGTCTCC	358		
Db	153284	GATGGCCACATGAAGCTGGTCGACGAAGCCAAAGCCCGCCCGATGTGTCGTCGTCAGT	153225		
QY	359	ATCTACGTCAACCCGGGCCAATTTCGCCCCCAAGGAGACCTCTCCACCTACCCCTCCGAC	418		
Db	153224	ATTTTCGTTAACCCGATGCAGTTCGACCGCCCGGAGACCTGGCTCGGTACCCACGCACC	153165		
QY	419	TTGACGGCGACGTAAAGAAACTCGCGTCCGTTCCCGCGCGCTCGACGTCGTTTTCAT	478		
Db	153164	TTGCAGGAAGACTGCGAAAAAGCT-----	153142		
QY	479	CCCCGTAACTTGTACGATTACGGGAAGAACGGTGGTGGTGACGTGGCAGAGCGTGGTGA	538		
Db	153141	-----GACAAGCGCAAAAGTGGATCTGGTTTCGGCCCTCG	153105		
QY	539	ATGGTGTCTGTCGTTGAGAGTGGGTCCGGGCACGAAAAGTTGGGTGAGGGTTCAGAAAGCTG	598		
Db	153104	GTAAAAGAGATCTACCCGAACGGTACTGAAACCCATACCTTACGTTGACGTTCCAGGGCTT	153045		
QY	599	GAATTGGGGCTGTGTGGGAAGAGCAGGCCCGCTTTCTTTCAGAGGGGTGGGACTGTGGTG	658		
Db	153044	TCGACCATGTGGAAGCGCCAGCCCGCGGGGCATTTTCGCGCGCTGTGACTATCGTC	152985		
QY	659	ACGAAGTTGTTTAATATTTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAG	718		
Db	152984	AGCAAGCTGTTTAACTGGTCCAGCCGGACATCGCTGCTTCGGTGAAGAAGGACTTTCAG	152925		
QY	719	CAGTGGCGGCTTATTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAAGTGATAGGT	778		
Db	152924	CAACTGGCGCTGATCCGCAAAATGGTTGCTGATATGGGCTTTGATATTGAGATTGTCGGT	152865		
QY	779	GCTGAAATAACACGCTGATAATGATGGCCCTGGCAATGAGTTCACGTAATGTGCACCTTTCA	838		
Db	152864	GTGCCAATTATGCGCGCCAAAGACGGTCTGGCAGCTGAGTTCGCCGTAAACGGTTATCTGACG	152805		
QY	839	CCTGAAGAGAGGGGAAAAGGCATATCAATAAATAAAATCATTGTTAAGAGCTAAATCAGCA	898		
Db	152804	GCAGAACAAACGCAAAATTTGCGCCCGTCTGTACAAAAGTTTAAAGTTGATTCGCGACAAA	152745		
QY	899	GCAGGAGATGGTCAGGTGCATTGTGTGAGAGATTGACAAAATTTGGTCAATCCAAAAGTTACT	958		
Db	152744	TTGCAGGCTGGCGAACCGGATCTCGATGAATATTGCTATTGCGGSCAAGAAGTGAAT	152685		
QY	959	GATGCTGTGGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAAATTTGGAGAAAAGTG	1018		
Db	152684	GAAAAGGCTTCCGCGCCGATGATATTTCAGATTCCGATGCCGACACATTGCTGGAAGTC	152625		
QY	1019	GAACAGATCAAGAGTCCTGTCGTTCTCTGTTGTCGATGGTTTGGCAAAAGTCAGGCTT	1078		
Db	152624	TCTGAAACCAAGCAACCGGCGAGTAATTCTGTAGCCGCTGGCTTGGCGATGCTCGCCTG	152565		
QY	1079	ATAGACAAATCGAAATCAACTTGTCAATGAATGTTTGAATCTAA	1122		
Db	152564	ATCACAACAAATAATGGTCGAGCTGGCGTAATACTTATACTCTAA	152521		



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 21:18:00 ; Search time 3107 Seconds  
(without alignments)  
10294.397 Million cell updates/sec

Title: US-10-033-269-8  
Perfect score: 1316  
Sequence: 1 tcggcacgaggtttctcag.....tggcattgctattgctagg 1316

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				EST:*
1:	em_estba:	*		
2:	em_esthum:	*		
3:	em_estin:	*		
4:	em_estmu:	*		
5:	em_estov:	*		
6:	em_estpl:	*		
7:	em_estro:	*		
8:	em_hic:	*		
9:	gb_est1:	*		
10:	gb_est2:	*		
11:	gb_hic:	*		
12:	gb_est3:	*		
13:	gb_est4:	*		
14:	gb_est5:	*		
15:	em_estfun:	*		
16:	em_estom:	*		
17:	em_gss_hum:	*		
18:	em_gss_inv:	*		
19:	em_gss_pln:	*		
20:	em_gss_vrt:	*		
21:	em_gss_fun:	*		
22:	em_gss_mam:	*		
23:	em_gss_mus:	*		
24:	em_gss_pro:	*		
25:	em_gss_rod:	*		
26:	em_gss_phg:	*		
27:	em_gss_vrl:	*		
28:	gb_gssi:	*		
29:	gb_gss2:	*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	629	47.8	629	13	BU965091
2	488.8	37.1	573	13	BQ453747
3	416	31.6	571	12	BM731347
4	346	26.3	1784	11	AY110983

5	309.8	23.5	541	13	BU034064
6	303.2	23.0	676	10	BG098937
7	288.2	21.9	576	13	BU030691
8	283.8	21.6	625	13	BU031672
9	283	21.5	503	13	BQ976077
10	276.6	21.0	531	9	AI730340
11	259.2	19.7	616	14	CB629807
12	254.4	19.3	545	9	AU290923
13	251.4	19.1	746	14	CB629808
14	247	18.8	875	29	BZ547854
15	244	18.5	838	10	BF064760
16	242.2	18.4	651	14	CB879036
17	241.4	18.3	573	9	AU286177
18	230.4	17.5	625	13	BQ768349
19	219	16.6	610	28	BH505138
20	218.2	16.6	575	29	BZ408845
21	203.8	15.5	425	9	AW719829
22	191.8	14.6	577	14	CB091543
23	190	14.4	488	9	AL829512
24	162.6	12.4	455	9	AU286178
25	160	12.2	606	9	AI998016
26	153.4	11.7	450	14	CA687321
27	144	10.9	404	14	CA725201
28	141.4	10.7	215	10	BF153835
29	138.8	10.5	779	29	BZ408852
30	127	9.7	225	13	BQ765881
31	124.2	9.4	868	10	BF267302
32	118	9.0	571	13	BU049399
33	114.8	8.7	545	12	BM276865
34	110.2	8.4	516	12	BM419808
35	106.6	8.1	518	12	BM276771
36	105.6	8.0	495	29	AG222917
37	104.4	7.9	439	12	BI203409
38	101.6	7.7	563	14	CA758922
39	100	7.6	564	10	BE040497
40	96.4	7.3	1065	29	CNS078AS
41	95	7.2	725	29	BZ560200
42	91.8	7.0	988	29	BZ573502
43	88.6	6.7	477	12	BM259201
44	83.4	6.3	581	13	BQ590858
45	83	6.3	451	9	AU173408

ALIGNMENTS

RESULT 1	BU965091	629 bp	linear	EST 21-OCT-2002
LOCUS	sat06b11.y1	Gm-cl036	Glycine max	cdNA clone SOYBEAN CLONE ID:
DEFINITION	Gm-cl036-12765	5'	similar to TR:024035	O24035
	PANTOATE--BETA-ALANINE LIGASE	;	mrna sequence.	
ACCESSION	BU965091			
VERSION	BU965091.1	GI:24205838		
KEYWORDS	EST.			
SOURCE	Glycine max (soybean)			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.			
REFERENCE	1	(bases 1 to 629)		
AUTHORS	Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.			
	Public Soybean EST Project			
TITLE	Unpublished			
JOURNAL	Contact: Shoemaker R/Public Soybean EST Project			
COMMENT	Public Soybean EST Project			
	Washington University School of Medicine			

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact: ccu@resgen.com web site:  
www.resgen.com

Seq primer: -40RP from Gibco  
High quality sequence stop: 446.

FEATURES

source

1. .629  
/organism="Glycine max"  
/mol\_type="mRNA"  
/db\_xref="taxon:3847"  
/clone="SOYBEAN CLONE ID: Gm-cl036-12765"  
/tissue\_type="somatic embryos cultured on MSD 20"  
/lab\_host="DH10B"  
/clone\_lib="Gm-cl036"

/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This  
cDNA library was constructed from mRNA isolated from  
somatic embryos (age ranging from 2 months to 9 months)  
cultured on MSD 20. The library was prepared using the  
Life Technologies pSuperScript cDNA library construction  
kit. Complementary DNA was synthesized from mRNA using a  
poly (dT) sequence with a NotI restriction site. SalI  
linkers adapters were ligated to the blunt-ended cDNA  
fragments followed by NotI digestion. The cDNA fragments  
were directionally cloned into the NotI-SalI restriction  
site of the pSPORT1 vector. The ligated cDNA fragments  
were transformed into E.coli Electromax DH10B host cells.  
This library was constructed in the laboratory of Dr. Lila  
Vodkin by Anu Khanna at the University of Illinois at  
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

BASE COUNT 124 a 197 c 173 g 135 t

Query Match

Best Local Similarity 47.8%; Score 629; DB 13; Length 629;

Matches 629; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 ACACACCAAACGTAACCTTCAACCCCAAGAGTCAATAAATTTGCTTAACACCTACCTCT 98  
Db 1 ACACACCAAACGTAACCTTCAACCCCAAGAGTCAATAAATTTGCTTAACACCTACCTCT 60  
QY 99 TCAACCAACCGGTTTCGCGCCATCAGCTTCGTCCTTTTACCTATTACACTCCTCAACGAAC 158  
Db 61 TCAACCAACCGGTTTCGCGCCATCAGCTTCGTCCTTTTACCTATTACACTCCTCAACGAAC 120  
QY 159 TCCCACTGTTTCCAATATTATTACATATGGCTCCAGCCCAAGGTGATCTCCGACAGG 218  
Db 121 TCCCACTGTTTCCAATATTATTACATATGGCTCCAGCCCAAGGTGATCTCCGACAGG 180  
QY 219 CCTCGATCGGAGCTGGTCGGCTCGATCGCGGCCCCAGGCAAGCTCATTGGGTGGTCC 278  
Db 181 CCTCGATCGGAGCTGGTCGGCTCGATCGCGGCCCCAGGCAAGCTCATTGGGTGGTCC 240  
QY 279 CCACATCGGCTTCTTCCACCGGGCCACCTCTCGCTCGTGCCAGGCCCGCCCAACTCT 338  
Db 241 CCACATCGGCTTCTTCCACCGGGCCACCTCTCGCTCGTGCCAGGCCCGCCCAACTCT 300  
QY 339 CCGACGTCGTCGCGCTCTCCATCTACGTCAACCCGGGCCAATTCGCCCCCAGGAGACC 398  
Db 301 CCGACGTCGTCGCGCTCTCCATCTACGTCAACCCGGGCCAATTCGCCCCCAGGAGACC 360  
QY 399 TCTCCACTACCCCTCCGACTTCGACGGCGACGTAAAGAACTCGGTCGGTCCGTTCCCGCG 458  
Db 361 TCTCCACTACCCCTCCGACTTCGACGGCGACGTAAAGAACTCGGTCGGTCCGTTCCCGCG 420  
QY 459 GCGTCGACGTCGTTTTCATCCCGTAACCTTGATCGATTACGGGAAGAACGGTGGTGGTG 518  
Db 421 GCGTCGACGTCGTTTTCATCCCGTAACCTTGATCGATTACGGGAAGAACGGTGGTGGTG 480

QY 519 ACGTGGCAGAGGCTGGTGAATGTCGTCGTCGTTGAGAGTGGTCCGGGCACGAAAGTT 578  
Db 481 ACGTGGCAGAGGCTGGTGAATGTCGTCGTCGTTGAGAGTGGTCCGGGCACGAAAGTT 540  
QY 579 GGGTGAGGTTGAGAAGCTGGAATTGGGGCTGTGTGGGAAGAGCAGCCCGTTTCTTCA 638  
Db 541 GGGTGAGGTTGAGAAGCTGGAATTGGGGCTGTGTGGGAAGAGCAGCCCGTTTCTTCA 600  
QY 639 GAGGGGTGGCGACTGTGGTGACGAAAGTTG 667  
Db 601 GAGGGGTGGCGACTGTGGTGACGAAAGTTG 629

RESULT 2

BQ453747/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .573  
/organism="Glycine max"  
/mol\_type="mRNA"  
/db\_xref="taxon:3847"  
/clone="SOYBEAN CLONE ID: Gm-cl081-2478"  
/tissue\_type="Roots of 7 day old 'Bragg' seedlings"  
/dev\_stage="7 days old"  
/lab\_host="DH10B"  
/clone\_lib="Gm-cl081"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:  
XhoI; The mRNA was isolated from roots of 7 day old  
'Bragg' seedlings that were mock-infected 48 hours prior  
to harvest. Dr. Gary Stacey generously donated the  
tissue. The roots were flash-frozen in liquid nitrogen.  
Stratagene's cDNA Synthesis Kit (catalog number 200401)  
was used to synthesize the cDNA. First-strand synthesis  
was performed with 5-methyl dCTP, hence the ligated cDNA  
was hemimethylated. A modification of Stratagene's  
first-strand synthesis primer was used. An 'anchor'  
nucleotide (V=A, C, or G) was added to the 3' end of the  
primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to anchor

BQ453747 573 bp mRNA linear EST 29-MAY-2002  
sao86d03.y1 Gm-cl081 Glycine max cDNA clone SOYBEAN CLONE ID:  
Gm-cl081-2478 5' similar to TR:O24035 O24035 PANTOATE--BETA-ALANINE  
LIGASE ;, mRNA sequence.

BQ453747 GI:21256859

EST.

Glycine max (soybean)

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

1 (bases 1 to 573)

Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna

, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers

, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk

, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

, R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact: ccu@resgen.com web site:

www.resgen.com

Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 474.

Location/Qualifiers

the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first- strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using Sephacryl S-500 High Resolution (Pharmacia Biotech) in a 2-mm diameter column and a bed volume of approximately 1ml. The column eluent was precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). This library was constructed in the laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at Northern Arizona University."

BASE COUNT	175 a	119 c	93 g	185 t	1 others
ORIGIN					
Query Match	37.1%; Score 488.8; DB 13; Length 573;				
Best Local Similarity	99.4%; Pred. No. 7.7e-126;				
Matches 490; Conservative	0; Mismatches	3; Indels	0; Gaps	0;	
QY	824	AATGTGCACCTTTACCTGAAGAGAGGGGAAAAGGCACCTATCAATAAAATCAATCATTTGTTA	883		
Db	573	AATGTGCACCTTTACCTGAAGAGAGGGGAAAAGGCACCTATCAATAAAATCAATCATTTGTTA	514		
QY	884	AGAGCTAAATCAGCAGCAGGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTC	943		
Db	513	AGAGCTAAATCAGCAGCAGGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTC	454		
QY	944	ATCCAAAGTGTACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTGATCAAAAAT	1003		
Db	453	ATCCAAAGTGTACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTGATCAAAAAT	394		
QY	1004	AATTTGGAGAAAGTGAACAGATCAAGAGTCCTGTCGTTCTTCTGTGTTGCTGTCATGGTTT	1063		
Db	393	AATTTGGAGAAAGTGAACAGATCAAGAGTCCTGTCGTTCTTCTGTGTTGCTGTCATGGTTT	334		
QY	1064	GGCAAAGTCAGGCTTATAGACAAACATGGAATCAACTTGTCAATGAATGTTTGATCTAAC	1123		
Db	333	GGCAAAGTCAGGCTTATAGACAAACATGGAATCAACTTGTCAATGAATGTTTGATCTAAC	274		
QY	1124	CTTCTGTATCTCAAAACATGGGCCACATGCTTAATTAATAGTTTCGGGCCACGTGCTTAAC	1183		
Db	273	CTTCTGTATCTCAAAACATGGGCCACATGCTTAATTAATAGTTTCGGGCCACGTGCTTAAC	214		
QY	1184	AATCTAACAGTTTCATGGTTATAGTCATGACAAATTTTTTTTCTGCCAGCCATACATGAT	1243		
Db	213	AATCTAACAGTTTCATGGTTATAGTCATGACAAATTTTTTTTCTGCCAGCCATACATGAT	154		
QY	1244	TACTTGTAGATGCATTTTACCGCATCATATAAATTTCTATGAGAGCTGCTACCGTGGCATT	1303		
Db	153	TACTTGTAGATGCATTTTACCGCATCATATAAATTTCTATGAGAGCTGCTACCGTGGCATT	94		
QY	1304	GCTATTTGCTAGG	1316		
Db	93	GCTATTTGCTAGG	81		
RESULT 3					
BM731347					
LOCUS					
DEFINITION	BM731347 sal70g09.y1 Gm-cl061 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-cl061-4553 5' similar to TR:024035 O24035 PANTOATE--BETA-ALANINE LIGASE ; mRNA sequence.				
ACCESSION	BM731347				
VERSION	BM731347.1 GI:19052680				
KEYWORDS	EST.				
SOURCE	Glycine max (soybean)				

ORGANISM	Glycine max				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.				
REFERENCE	1 (bases 1 to 571)				
AUTHORS	Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.				
TITLE	Public Soybean EST Project				
JOURNAL	Unpublished				
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com Seq primer: -40RP from Gibco High quality sequence stop: 426. Location/Qualifiers 1. 571				
FEATURES	source				
	/organism="Glycine max" /mol_type="mRNA" /db_xref="taxon:3847" /clone="SOYBEAN CLONE ID: Gm-cl061-4553" /tissue_type="mature flowers of field grown plants" /lab_host="DH10B" /clone_lib="Gm-cl061" /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from mature flowers of field grown plants for the cultivar Raiden. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."				
BASE COUNT	167 a	93 c	126 g	185 t	
ORIGIN					
Query Match	31.6%; Score 416; DB 12; Length 571;				
Best Local Similarity	100.0%; Pred. No. 2.2e-105;				
Matches	416; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	901	AGGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTGTACTGA	960		
Db	1	AGGAGATGGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTGTACTGA	60		
QY	961	TGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGA	1020		
Db	61	TGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGA	120		
QY	1021	ACAGATCAAGAGTCCTGTCCTTCTGTGTTGCTGCATGGTTTGGCAAAGTCAGGCTTAT	1080		
Db	121	ACAGATCAAGAGTCCTGTCCTTCTGTGTTGCTGCATGGTTTGGCAAAGTCAGGCTTAT	180		
QY	1081	AGACAAACATGGAAATCAACTTGTCAATGAATGTTTGAATCTAACCTTCTGTCTCATCAAAC	1140		
Db	181	AGACAAACATGGAAATCAACTTGTCAATGAATGTTTGAATCTAACCTTCTGTCTCATCAAAC	240		
QY	1141	ATGGGCCACATGCTTAATAATAGTTTCGGGCCACGTGCTTAACAATTTCAACAGTTCATG	1200		



[illegible]

TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project  
http://compgenomics.ucdavis.edu/  
JOURNAL Unpublished  
COMMENT Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@atgc.org [michelmore@vegmil.ucdavis.edu]  
belongs to contig QH\_CA\_Contig3643, see http://cgpdb.ucdavis.edu/  
for details.  
Plate: QHJ2 row: P column: 22.  
Location/Qualifiers  
1. .541  
/organism="Helianthus annuus"  
/mol\_type="mRNA"  
/cultivar="RHA280"  
/db\_xref="taxon:4232"  
/clone="QHJ2P22"  
/lab\_host="E.coli"  
/clone\_lib="QH\_EFGHJ sunflower RHA280"  
/note="Vector: pBRCDNASfiAB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/  
TAG\_LIB=QH\_EFGHJ sunflower RHA280  
TAG\_TISSUE=germinating seeds  
TAG\_SEQ=TCTGTGCGGG"  
BASE COUNT 152 a 71 c 176 g 142 t  
ORIGIN  
Query Match 23.5%; Score 309.8; DB 13; Length 541;  
Best Local Similarity 73.6%; Pred. No. 1.5e-75;  
Matches 395; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 568 GCACGAAAGTTGGGTGAGGGTTGAGAAAGCTGGAATTGGGGCTGTGTGGGAAGACGAGGCC 627  
Db 1 GCATGAGACATGGGTTAGAGTTGAGAAGCTGGAGAAGGGATGTGTGGGAAGAGTAGGCC 60  
QY 628 CGTCTTCTCAGAGGGGTGGCGACTGTGGTGACGAAGTTGTTTAATATTGTGAGCCAGA 687  
Db 61 TGTGTTTTTTAGAGGGGTGGCTACTGTGTGGCTAAGTTGTTTAATATTGTGAACCGGA 120  
QY 688 TGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGGCGGCTTATTTCAGAGGATGGTTTCG 747  
Db 121 TGTGCGGTTTTTGGGAAGAAGGATTATCAGCAGTGGCGGTTATTTCGCGGATGGTTTCG 180  
QY 748 AGATCTTGATTTTCCATAAAAGTGATAGGTGCTGAAATAACACGTCGATAATGATGGCCT 807  
Db 181 AGATCTTGATTTTGTAAAGTGATTGGTTCTGAGTTGAAGCGGAGAAAGATGGCCT 240  
QY 808 GGCAATGAGTTCACGTAATGTGCACCTTTCACCTGAAGAGAGGGGAAAAGGCACATATCAAT 867  
Db 241 GGCAATGAGTCCCGCAACGTGCATCTTTTCAACAAGAGAGGGGAAACAGCCCTTGTGCAT 300  
QY 868 AAATAAATCATGTTAAGAGCTAAATCAGCAGCAGGAGATGGTCAGGTGCATTGTGAGAA 927  
Db 301 AAGCCGGGCTTTATTGAAGCCAAATCAGCTGCCAAAAGGGCAATAGCAATTGCAGTAA 360  
QY 928 GTTGACAAATTTGTTTCATCCAAAGTGTACTGATGCTGGTGGGAAGGATCGAATTATGCTGA 987  
Db 361 GTTGAGAAGATCAGTCATCGAAAGTACAGAAACAGCCCGGTGGAAAAATAGATTATGCTGA 420  
QY 988 GATTGTTGATCAAAATAATTTGGAGAAAGTGGAAACAGATCAAGAGTCTCTGTCGTTCTTG 1047  
Db 421 GATTGTGATCAAGAAAGTCTAGAAACGGTTGAAGAGATCAAGGATGCGACTTGTGGTGTG 480

QY 1048 TGTGTGCTGCATGGTTTGGCAAAGTCAGGCTTATAGACAACATGGAATCAACTTGTG 1104  
Db 481 TGCTGCAGCATGGTTTGGGAACGTGAGGTTGATCGACAACATTGAAATTGATCGGTC 537  
RESULT 6  
BG098937/c  
LOCUS BG098937 676 bp mRNA linear EST 07-MAR-2003  
DEFINITION EST463456 sprouting eyes/shoots Solanum tuberosum cDNA clone  
CSTC8O20 3' sequence, mRNA sequence.  
ACCESSION BG098937  
VERSION BG098937.1 GI:12588972  
KEYWORDS EST.  
SOURCE Solanum tuberosum (potato)  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum.  
REFERENCE 1 (bases 1 to 676)  
AUTHORS van der Hoeven,R.S., Bezzerides,J., Cho,J., Utterback,T., Hansen  
,C.L., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker  
,B.  
TITLE Generation of ESTs from potato sprouting eyes/shoots  
JOURNAL Unpublished  
COMMENT Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato-array@tigr.org  
This clone can be obtained from the University of Arizona Genomics  
Institute. Orders can be made through URL:  
http://genome.arizona.edu/orders/.  
FEATURES  
Location/Qualifiers  
1. .676  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="CSTC8O20"  
/tissue\_type="sprouting tubers"  
/dev\_stage="12-14 weeks post harvest"  
/lab\_host="SOLR"  
/clone\_lib="sprouting eyes/shoots"  
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
XhoI; Various sizes of sprouting eyes (2 mm to 15 mm) were  
taken from tubers. The tubers were incubated at 26C in  
the dark for 2-3 weeks prior to sprouting. The eyes were  
frozen in liquid nitrogen immediately upon removal from  
tubers."  
BASE COUNT 205 a 173 c 92 g 206 t  
ORIGIN  
Query Match 23.0%; Score 303.2; DB 10; Length 676;  
Best Local Similarity 75.8%; Pred. No. 1.2e-73;  
Matches 388; Conservative 0; Mismatches 123; Indels 1; Gaps 1;  
QY 597 TGGAAATTGGGCTGTGTGGGAAGACAGGCCCGCTTTTCTTCAGAGGGTGGCGACTGTGG 656  
Db 675 TGGAAAAGGGATTGTGTGGGAATAGTAGCCCTGTTTTCTTTAAAGGGTTGCTACTGTTG 616  
QY 657 TGACGAAGTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATC 716  
Db 615 TTACCAAGTTGTTCAATATAGTTGAGCCTGATGTTGTTCTTTTGGTAAGAGGATTATC 556  
QY 717 AGCAGTGGCGGCTTATTTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAGTGATAG 776  
Db 555 AGCAATGGAGGATTATACAGAGAATGGTCAGAGATCTTGATTTGGGATAAAGTGATTG 496  
QY 777 GTGCTGAAATAACACGTCGATATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTT 836  
Db 495 GTTCTGAAATAGTACGAGAGCATGATGGCCCTTGCAATGAGTTCCTCGTAATGTGAAACTTT 436  
QY 837 CACCTGAAGAGAGGGAAAAGGCACTATCAATAATAATCAATTGTTAAGAGCTAAATCAG 896

Db	435	CACCTGCGGACAGACAAAGGGCTTTATCAATTAGTCGTGGCTGTCTAGAGCAACAAGTTG	376
QY	897	CAGCAGGAGATGGTTCAGGTGCATTTGTGAGAAGTTGCACAAATTTGGTTCATCCAAAGTGTTA	956
Db	375	AAGCAGAAAAAGGTCAGGTCAACTGCAGGGAGCTGATAAAATACTGCCATTCTTACCATAT	316
QY	957	CTGATGCTGGTGGAAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAG	1016
Db	315	CTGAAGCTGGTGAACCGGTGATTATGCTGAGATTGTGGATCAAGAAAGTTTAGAGCCAG	256
QY	1017	TGGAACAGATCAAGAGAGTCCTGTCGTTCTTCTGTGTTG-CTGCATGGTTTGGCAAAGTCAGG	1075
Db	255	TGGAACAATCAAGAGAGACCAGTTGTATTTTGTGTAGCCAGTATGGTTTGGAAAGTCAGG	196
QY	1076	CTTATAGACAACATGGAATCAACTTGTCAAT	1107
Db	195	CTTATTGACAACATGGAATAGATGTATAAT	164
RESULT 7			
BU030691			
LOCUS			
DEFINITION	BU030691	576 bp mRNA linear EST 23-AUG-2002	
	QHJ16C05.Yg.ab1 QH_EFGHJ sunflower RHA280	Helianthus annuus cDNA	
ACCESSION	clone QHJ16C05,	mRNA sequence.	
VERSION	BU030691		
KEYWORDS	BU030691.1	GI:22466211	
SOURCE			
ORGANISM	Helianthus annuus (common sunflower)		
	Helianthus annuus		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	asterids; campanulids; Asterales; Asteraceae; Asteroideae;		
	Heliantheae; Helianthus.		
REFERENCE	1 (bases 1 to 576)		
AUTHORS	Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,		
	Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison		
	,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,		
	Church,S., Jackson,L. and Bradford,K.		
TITLE	Lettuce and Sunflower ESTs from the Compositae Genome Project		
JOURNAL	http://compgenomics.ucdavis.edu/		
COMMENT	Unpublished		
	Contact: Alexander Kozik [R.W.Michelmore]		
	Department of Vegetable Crops, R.W.Michelmore Lab		
	University of California at Davis (UCD)		
	Asmundson Hall, UCD, Davis, CA 95616, USA		
	Tel: 1-(530)-742-1742		
	Fax: 1-(530)-752-9659		
	Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]		
	belongs to contig QH_CA_Contig3643, see http://cgdb.ucdavis.edu/		
	for details.		
FEATURES	Plate: QHJ16	row: C	column: 05.
source	Location/Qualifiers		
	1..576		
	/organism="Helianthus annuus"		
	/mol_type="mRNA"		
	/cultivar="RHA280"		
	/db_xref="taxon:4232"		
	/clone="QHJ16C05"		
	/lab_host="E.coli"		
	/clone_lib="QH_EFGHJ sunflower RHA280"		
	/note="Vector: pBRCDNAsfIAB; The library was constructed		
	from 11 different sources of RNA from a single genotype.		
	Separate cDNAs were generated using primers that		
	incorporated unique 5' and 3' tags to distinguish each		
	source of RNA. cDNAs were then pooled, size-fractionated,		
	directionally cloned into a custom medium-copy vector and		
	transformations made with four size classes to minimize		
	size bias. Details of each source of RNA and library		
	construction can be obtained at http://cgdb.ucdavis.edu/		
	TAG_LIB=QH_EFGHJ sunflower RHA280		
	TAG_TISSUE=germinating seeds		
	TAG_SEQ=TCTGTGCGGG		
BASE COUNT	163 a	76 c	183 g 154 t



for details. Plate: QHJ19 row: C column: 05. Location/Qualifiers 1. .625 /organism="Helianthus annuus" /mol_type="mRNA" /cultivar="RHA280" /db_xref="taxon:4232" /clone="QHJ19C05" /lab_host="E.coli" /clone_lib="QH_EFGHJ sunflower RHA280" /note="Vector: pBRCDNASfiAB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_LIB=QH_EFGHJ sunflower RHA280 TAG_TISSUE=germinating seeds TAG_SEQ=TCTGTGCGGG"									
BASE COUNT		179 a	76 c	188 g	179 t	3 others			
ORIGIN									
Query Match 21.6%; Score 283.8; DB 13; Length 625; Best Local Similarity 72.6%; Pred. No. 3.2e-68; Matches 392; Conservative 0; Mismatches 145; Indels 3; Gaps 2;									
QY	568	GCACGAAAGTTGGGTGAGGGTTGAGAAAGCTGGAATTTGGGGCTGTGT-GGGAAAGACAGGC	626						
Db	1	GCATGAGACATGGGTTAGAGTTGAGAAAGCTGGAGAAAGCGATGTGTGGGGAAGATANGC	60						
QY	627	CC--GTTTCTTCAGAGGGGTGGCGACTGTGTGACAAAGTTGTTTAATATTGTGGAGCC	684						
Db	61	CCTTGTGTTTTTGTAGAGGGTGGCTACTGTTGTGGCTAAGTTGTTTAATATTGTGAACC	120						
QY	685	AGATGTGGCTGTGTTCCGTTAAGAAGGATTATCAGCAGTGGCGGCTTATTTCAGAGGATGGT	744						
Db	121	GGATGTTGCGGTTTTTGGGAAGAAGGATTATCAGCAGTGGCGGTTATTTCGACGGATGGT	180						
QY	745	TCGAGATCTTGATTTTTCCATAAAAGTGATAGGTGCTGAAATAACACGCTGATAATGATGG	804						
Db	181	TCGAGATCTTGATTTTGTCTGTAGAGTGATTTGTTCTGAGTTGAAGCGGAGAGAAGATGG	240						
QY	805	CCTGGCAATGAGTTTCACGTAATGTGCACCTTTACCTGAAGAGAGGGGAAAAGGCATATC	864						
Db	241	CCTGGCAATGAGCTCCCGCAACGTGCATCTTTCAACAAGAGAGGGGAACAGGCCTTGTC	300						
QY	865	AATAAATAAATCATTTGTTAAGAGCTAAATCAGCAGCAGGAGATGGTCAGGTGCATTGTGA	924						
Db	301	GATAAGCCGGGCTTTATTGTAAGCCANATCAGCTGCCAAAAAGGCAATAGCAATTGCAG	360						
QY	925	GAAGTTGACAAATTTGGTTCATCCAAAGTGTACTGATGCTGGTGGAAAGGATCGATTATGC	984						
Db	361	TAAGTTGAGAAGATCAGTCGATCGAAAGTATAGAAGCAGCCGGTGGAAAAATAGATTATGT	420						
QY	985	TGAGATTGTTGATCAAAATAATTGGAGAAAGTGGAAACAGATCAAGAGTCTCTGTCGCTT	1044						
Db	421	TGAGATTGTGGATCAAGAAAGTCTAGAAAACGGTTGAAGAGATCAAGGATGAGTTGTGGT	480						
QY	1045	CTGTGTTGCTGCATGGTTTGGCAAAGTCAGGCTTATAGACAACATGGAATCAACTTGTTC	1104						
Db	481	GTGTGTTGCAGCATGGTTTGGGAAGGTGAGTTGATCGANCACATTGAAATTGATGCGTC	540						
RESULT 9 BQ976077 LOCUS DEFINITION QH120G07.yg.ab1 QH ABCDI sunflower RHA801 Helianthus annuus cDNA clone QH120G07, mRNA sequence. ACCESSION BQ976077									

BQ976077.1 GI:22393600 EST. Helianthus annuus (common sunflower) Helianthus annuus Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Asteroidae; Heliantheae; Helianthus. 1 (bases 1 to 503) Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Riesenberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K. Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/ Unpublished Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore Lab University of California at Davis (UCD) Asmundson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu] belongs to contig QH_CA_Contig3643, see http://cgpdb.ucdavis.edu/ for details. Plate: QH120 row: G column: 07. Location/Qualifiers 1. .503 /organism="Helianthus annuus" /mol_type="mRNA" /cultivar="RHA801" /db_xref="taxon:4232" /clone="QH120G07" /lab_host="E.coli" /clone_lib="QH ABCDI sunflower RHA801" /note="Vector: pBRCDNASfiAB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_LIB=QH ABCDI sunflower RHA801 TAG_TISSUE=germinating seeds TAG_SEQ=TCTGTGCGGG"									
BASE COUNT		137 a	61 c	166 g	139 t				
ORIGIN									
Query Match 21.5%; Score 283; DB 13; Length 503; Best Local Similarity 72.9%; Pred. No. 4.8e-68; Matches 364; Conservative 0; Mismatches 135; Indels 0; Gaps 0;									
QY	566	GGGCACGAAAGTTGGGTGAGGGTTGAGAAGCTGGAATTGGGGCTGTGTGGGAAGAGCAGG	625						
Db	5	GGGCATGAGACATGGGTTAGAGTTGAGAAGCTGGAGAAAGGATGTGTGGGAAGAGTAGG	64						
QY	626	CCCGTTTTCTTCAGAGGGGTGGCGACTGTGGTGACGAAGTTGTTTAATATTGTGGAGCCA	685						
Db	65	CCTGTGTTTTTATAGGGGTGGCTACTGTTGTGGCTAAGTTGTTTAATATTGTGAACCG	124						
QY	686	GATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGGCGGCTTATTCAGAGGATGGTT	745						
Db	125	GATGTTGCGGTTTTTGGGAAGAGGAGATTATCAGCAATGGCGGGTTATTCGTCGGATGGTT	184						
QY	746	CGAGATCTTGATTTTTCCATAAAAGTATAGGTGCTGAAATAACACGTGATAATGATGGC	805						
Db	185	CGAGATCTTGATTTTGTCTGTTAAAGTGATTGTTCTGAGTTGATCGAGAAGAGATGGC	244						
QY	806	CTGGCAATCAGTTTCACGTAATGTGCACCTTTTCACCTGAAGAGAGGGGAAAAGGCATATCA	865						
Db	245	CTGGCAATGAGCTCCCGCAACGTCATCTTTCAACAAGAGAGGGGAACAGGCCTTGTCG	304						

QY 866 ATAAATAAATCATTTGTTAAGAGCTAAATCAGCAGCAGGAGATGGTCAGGTGCATTGTGAG 925  
|||||  
Db 305 ATAAGCGGGCTTTATTTGAAGCCAAATCAGCTGCCAAAAGGGCAATAACAATTGCAGT 364  
QY 926 AAGTTGACAAATTTGGTCATCCAAAGTGTTACTGATGCTGGTGGAGGATCGATTATGCT 985  
|||||  
Db 365 GAGTTGAGAAGATCAGTGATCGAAAGTATAGAACGACCGGTGGAAAAATAGATTATGTT 424  
QY 986 GAGATTGTTGATCAAAATAATTTGGAGAAAGTGGAACAGATCAAGAGTCCTGTCGCTTTC 1045  
|||||  
Db 425 GAGATTGTGATCAAGAGAGCTCAGAAATGTTGAAGAGACGAATGATGGAGTTGTGGTG 484  
QY 1046 TGTGTTGCTGCATGTTTG 1064  
|||||  
Db 485 TGTGTTGACGATGTTTG 503  
RESULT 10  
AI730340  
LOCUS  
DEFINITION  
AI730340 531 bp mRNA linear EST 11-JUN-1999  
BNLGH16682 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar  
to (Y10252) pantoate--beta-alanine ligase [Lotus japonicus], mRNA  
sequence.  
ACCESSION  
AI730340  
VERSION  
AI730340.1 GI:5049192  
KEYWORDS  
EST.  
SOURCE  
Gossypium hirsutum (upland cotton)  
ORGANISM  
Gossypium hirsutum  
REFERENCE  
1 (bases 1 to 531)  
AUTHORS  
Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.  
TITLE  
ESTs from developing cotton fiber  
JOURNAL  
Unpublished  
COMMENT  
Contact: Ben Burr  
Biology Department  
Brookhaven National Laboratory  
Upton, NY 11973, USA  
Tel: 516-344-3396  
Fax: 516-344-3407  
Email: burr@bnlxl.bnl.gov  
Seq primer: T3 Primer.  
Location/Qualifiers  
1. 531  
/organism="Gossypium hirsutum"  
/mol\_type="mRNA"  
/cultivar="Acala Maxxa"  
/db\_xref="taxon:3635"  
/tissue\_type="immature fiber"  
/dev\_stage="Six days post anthesis"  
/lab\_host="XLI-Blue"  
/clone\_lib="Six-day Cotton fiber"  
/note="Vector: pBluescript II KS+"  
BASE COUNT 163 a 73 c 157 g 136 t 2 others  
ORIGIN  
Query Match 21.0%; Score 276.6; DB 9; Length 531;  
Best Local Similarity 70.3%; Pred. No. 3.1e-66;  
Matches 369; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
QY 485 AACTGTGACGATTACGGGAAGAACGGTGGTGACGTGGCAGAGGCTGGTGGGAATGGTG 544  
|||||  
Db 6 AATAAGAAATGGAATGAGAATTTGGGTGCTGGTGGGATGGGAATGGGGTGGTTCA 65  
QY 545 TCGTGGTTGAGAGTGGGTCGGGCGACGAAAGTTGGGTGAGGGTTGAGAAGCTGGAATTG 604  
|||||  
Db 66 TGTGTGGAGGANAANGGAGTGGGCATGAACATGGGTGAGAGTTGAAAGATTGGAGAAA 125  
QY 605 GGGCTGTGTGGGAAGACGAGGCCCGTCTTTCTTCAGAGGGGTGGCGACTGTGGTGACGAAG 664  
|||||  
Db 126 GGATTGTGTGGGAAAAGTAGGCCCGTGTCTTCAAAGAGGTTGCCACCATTTGTACCAAG 185

QY 665 TTGTTTAATATTGTGAGCCAGATGTGGCTGTGTTCGGTAAAGAAGGATTATCAGCAGTGG 724  
|||||  
Db 186 TTGTTTCACATTGTGGAACCTGATGTTGCTGTCTTTGGTAAAGGATTATCAGCAGTGG 245  
QY 725 CGGCTTATTTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAGTGATAGGTGCTGAA 784  
|||||  
Db 246 AGGATAATAACAACGGATGGTTCGAGATCTTGATTTGCCCATTTGAAATTGTGGGTCTGAA 305  
QY 785 ATAACACGTGATAATCATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTCACCTGAA 844  
|||||  
Db 306 ATAACCCGTGATAGCGATGGCCTCGCGCTGAGTTCACGAAATGTGCATCTCTCCCTGAA 365  
QY 845 GAGAGGGAAAAGGCACTATCAATAAATAAATCAATTTTAAGAGCTAAATCAGCAGCAGGA 904  
|||||  
Db 366 GAAAGGCCAAAAGGCATTGTCTATTACCAGGTCAATGTCAAAGGCTAAACCTGCTGCTAAA 425  
QY 905 GATGTCAGGTGCATTGTGAGAAGTTGACAAAATTTGGTCATCCAAAGTGTTACTGATGCT 964  
|||||  
Db 426 GAAAGGCCAAGTTAAATGTAAAGAACTAAAGGATTCAGCTATCCTAACATAACTCAAGCT 485  
QY 965 GGTGGAAGGATCGATTATGCTGAGATTGTTGATCAAAAATAATTG 1009  
|||||  
Db 486 GGGGGAAGAATTGATTATGCTGAGATAGTTGATCAGGAGAAATTG 530  
RESULT 11  
CB629807  
LOCUS  
DEFINITION  
CB629807 616 bp mRNA linear EST 08-APR-2003  
OSIIEB06F09.f OSIIEb Oryza sativa (indica cultivar-group) cDNA  
clone OSIIEB06F09 5', mRNA sequence.  
ACCESSION  
CB629807  
VERSION  
CB629807.1 GI:29624796  
KEYWORDS  
EST.  
SOURCE  
Oryza sativa (indica cultivar-group)  
ORGANISM  
Oryza sativa (indica cultivar-group)  
REFERENCE  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 616)  
AUTHORS  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
TITLE  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
JOURNAL  
Unpublished  
COMMENT  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtcg  
BACKWARD: gga aac agc tat gac cat g  
Plate: 06 row: F column: 09  
Seq primer: gta aaa cga cgg cca gtcg.  
Location/Qualifiers  
1. 616  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR36"  
/db\_xref="taxon:39946"  
/clone="OSIIEB06F09"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSIIEb"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"  
BASE COUNT 111 a 187 c 182 g 135 t 1 others  
ORIGIN









Search completed: December 16, 2003, 00:07:42  
Job time : 3118 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
  
OM nucleic - nucleic search, using sw model  
Run on: December 15, 2003, 18:51:12 ; Search time 415 Seconds  
(without alignments)  
8560.146 Million cell updates/sec  
  
Title: US-10-033-269-8  
Perfect score: 1316  
Sequence: 1 tcggcacgaggctttctcag.....tggcattgctatttgctagg 1316  
  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
  
Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512  
  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 19Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*

2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*

3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*

4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*

5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*

6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*

7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*

8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*

9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*

10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*

11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*

12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*

13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*

14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*

15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*

16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*

17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*

18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*

19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*

20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*

21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*

22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*

23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*

24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	625	47.5	1338	20	AAZ20926	Nucleotide sequence
2	426.8	32.4	2184	21	AAC46075	Arabidopsis thaliana
3	422.2	32.1	1189	21	AAC41869	Arabidopsis thaliana
4	354	26.9	1264	20	AAZ20927	Nucleotide sequence
5	140.8	10.7	908	21	AAC52194	Arabidopsis thaliana
C 6	109	8.3	1672	20	AAH13694	Enterococcus faecalis
C 7	109	8.3	1672	24	ABS99489	Enterococcus faecalis
C 8	101.2	7.7	10982	22	AAS46264	DNA encoding novel protein

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

9	99.8	7.6	690	22	AAH53230	S. epidermidis ope
10	99.8	7.6	867	24	ABN92021	Staphylococcus epi
C 11	99.8	7.6	4103	22	AAH54816	S. epidermidis gen
C 12	99	7.5	4549	18	AAV74485	Staphylococcus aur
C 13	91.6	7.0	78845	21	AAA81463	N. meningitidis pa
14	91.6	7.0	349980	21	AAF21608	Neisseria meningit
15	91.6	7.0	1437668	21	AAA81490	N. meningitidis B
C 16	86.6	6.6	1132	22	AAH55061	S. epidermidis gen
17	83.6	6.4	546	24	ABK73744	Bacillus lichenifo
18	82.6	6.3	858	22	AAS00989	B. subtilis panC g
19	82.6	6.3	2363	22	AAS02314	B. subtilis panBCD
20	82.6	6.3	4450	22	AAS02327	Plasmid pAN441 for
21	82.6	6.3	7140	24	AAL41228	Pantothenate relat
22	82.6	6.3	7140	24	AAL48884	B subtilis PanB ge
C 23	82.6	6.3	7381	22	AAS02335	Plasmid pAN006 car
C 24	82.6	6.3	8503	22	AAS02334	Plasmid pAN004 car
C 25	82.6	6.3	10801	22	AAS02323	Plasmid pAN240 car
26	79.8	6.1	858	24	ABQ90275	M. capsulatus gene
27	79.2	6.0	777	25	ABZ39410	N. gonorrhoeae nuc
28	78	5.9	1193	25	ABA00659	Human ENZM-7 cDNA,
29	78	5.9	1446	23	AAS92821	DNA encoding novel
C 30	73.4	5.6	640681	24	ABA92787	Buchnera sp. genom
31	71.6	5.4	100848	22	AAF28552	Genomic fragment #
32	70.8	5.4	837	22	AAH65097	C glutamicum codin
33	70.8	5.4	960	22	AAF72059	Corynebacterium gl
34	70.8	5.4	2164	21	AAA71997	C. glutamicum panB
35	70.8	5.4	2164	21	AAA40273	C. glutamicum panB
36	70.8	5.4	2164	21	AAA40282	C. glutamicum panB
C 37	70.8	5.4	349980	22	AAH64966	C glutamicum codin
C 38	69.2	5.3	4403765	22	AAI99683	Mycobacterium tube
C 39	69.2	5.3	4411529	22	AAI99682	Mycobacterium tube
C 40	67.4	5.1	837	23	AAS88471	DNA encoding novel
41	67.4	5.1	2712	23	AAS86723	DNA encoding novel
C 42	67.4	5.1	2712	23	AAS89038	DNA encoding novel
43	67.4	5.1	2712	23	AAS89081	DNA encoding novel
44	67.4	5.1	2712	23	AAS90054	DNA encoding novel
45	67.4	5.1	3061	23	AAS89720	DNA encoding novel

ALIGNMENTS

RESULT 1	
AAZ20926	
ID	AAZ20926 standard; DNA; 1338 BP.
XX	
AC	AAZ20926;
XX	
DT	02-DEC-1999 (first entry)
XX	
DE	Nucleotide sequence of L. japonicus Pantothenate Synthetase (PS).
XX	
KW	pantothenate synthetase; cloning; herbicide; biosynthesis;
KW	Coenzyme A precursor; pantoate-beta-alanine ligase; ds.
XX	
OS	Lotus japonicus.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	20..946
FT	/*tag= a
FT	/product= Pantothenate Synthetase
XX	
PN	WO9942565-A1.
XX	
PD	26-AUG-1999.
XX	
PF	02-JUN-1998; 98WO-EP03261.
XX	
PR	31-MAY-1997; 97GB-0011163.
PR	27-JUN-1997; 97GB-0013477.
XX	
PA	(AGRE ) HOECHST-SCHERING AGREVO GMBH.
XX	



PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.



PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 32.4%; Score 426.8; DB 21; Length 2184;			
Best Local Similarity 64.9%; Pred. No. 8e-103;			
Matches 654; Conservative 0; Mismatches 342; Indels 12; Gaps 1;			
QY	122	ACGTTTCGTCCTTTACACTTATTACACTCCTCAGAAACTCCACTGTTTCCAATATTATTC	181
Db	964	ACATTAAATTCAGTTATCTCTCCACGAGAAACATTCTGCGAAGACAATCTCAAT	1023
QY	182	ACAATGGTCCAGCCCCAAGGGTGATCTCCGACAAAGCCCTCGATCGGAGCTGGTCGGC	241
Db	1024	TTTCCGACAAATGGAACCAAGAGTAATCAGAGACAAAGACTCTATGAGAAAATGGTCTCGA	1083
QY	242	TCGATCGGGGCCAGGGCAAGCTCAFTGGGTGGTCCCCACCATGGGCTTCCTCCACGGC	301
Db	1084	GCCATGAGATCACAAAGGCAAGACCATGGTTAGTACCAACAATGGTTACCTCCACGAA	1143
QY	302	GGCCACCTCTCGTCTGGCCCCAGGCCCGCCAACTCTCCGAGTCTGTCGCCGCTCTCCATC	361
Db	1144	GGTCACATTATCTCTCGTTCGTCATCACTAGCCCTCACTGACGTCACCGTCGTCCTCATC	1203
QY	362	TACGTCAACCCGGGCCAATTCCGCCCCACCGGAGGACCTCTCCACCTACCCCTCCGACTTC	421
Db	1204	TACGTCAATCCAGGCCAATTCTCTCCACTGAAGACCTCTCCACATACCCATCTGATTC	1263
QY	422	GACGGCAGCTAAAGAAACTCGGTCGCTCCGTTCCGGCGGCTGCGACGTCGTTTTCCATCCC	481
Db	1264	TCCGGCATCTCACTAAGCTCGCGGCTCTTCCGGTGTTAAAGTCGTCGTTTAAACCCA	1323
QY	482	CGTAACCTGTACGATTACGG-----GAAGAACGGTGGTGACGTGGCAGAG	529
Db	1324	AAAAACCTTATGATTACGGCGGTGAGACGAAGATAAACGACGGTGGTGGTAATGTT	1383
QY	530	GCTGSTGGAATGGTGTGTCGTCGTTGAGAGTGGTCCGGGCACGAAAGTTGGGTAGGGTT	589
Db	1384	GGGAGGTAGTGAGTTGTGTGGAGGAAGTGGTTAGGGCATGAGACTTGGATTAGGTT	1443
QY	590	GAGAAGCTGGAATTGGGGCTGTGTGGGAAGACGAGGCCCGCTTTCTTCAGAGGGGTGGC	649
Db	1444	GAGAGATTGGAGAAAGGTTTTTGTGGGAAGAGTAGGCCCTGTGTTCTTTAGAGGTGTGCT	1503
QY	650	ACTGTGCTGACGAAGTTGTTTAAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAG	709
Db	1504	ACTATTGTTACTAAGCTTTTAAATATTGTTGAGCCTGATGTGCTCTGTTTGGTAAGAAA	1563
QY	710	GATTATCAGCAGTGGCGGCTTATTTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAA	769
Db	1564	GATTATCAACAATGGAGGATTATACAGAGAATGGTTCGAGATCTTAATTTTGGTAATGAG	1623
QY	770	GTGATAGGTGCTGAAATAACACGFGATAATGATGGCCTGGCAATGAGTTCACGTAATGTG	829
Db	1624	ATTGTAGGATCAGATATAGCTAGAGAAAAGATGGACTTGGCATGAGTTCGAGAAATGTG	1683
QY	830	CACCTTTCACCTGAAGAGAGGGGAAAGGCACCTATCAATAATAAATCATTTGTTAAGAGCT	889
Db	1684	CGTCTTTCAGATGAAGAAAGGCAAGGGCTTTGTCTATAAGTAGGTCACCTGGCCATGGCT	1743
QY	890	AAATCAGCAGCAGGAGATGGTCAGGTGCATTTGTGAGAAGTTGACAAAATTTGGTCATCAA	949
Db	1744	AAAGCTTCTGTTGCAGAAGGGAACCAATTTGTGCAGAGCTTAAGGATATGATCAATCAA	1803
QY	950	AGTGTACTGATGCTGGTGGAAGGATCGATTATGCTGAGATGTTGATCAAAATAATTG	1009

Db	1804	CAAAGTTGTTGGATCTGCAGGACGAGTAGATTATATGCGAGATAGTTGACCAAGAAACTCTT	1863
QY	1010	GAGAAAGTGGAAACAGATCAAGAGTCCCTGTCGTCCTTCTGTGTTGCTGTCATGTTGGCAAA	1069
Db	1864	GAAGGAGTAGAAGAGATAAAGAGTGGGTAGTGATTTGTGTTGCTGCTTGGTTGGAACG	1923
QY	1070	GTCAGGCTTATAGACAACATGGAAATCAAACCTTGTCATGAATGTTTGA	1117
Db	1924	GTCAGGCTCATAGACACACATTGAGATCAATGTCTCTCTCTAGATCTCA	1971
RESULT 3			
AAC41869			
ID	AAC41869 standard; DNA; 1189 BP.		
XX	AAC41869;		
AC	AAC41869;		
XX	17-OCT-2000 (first entry)		
DT	17-OCT-2000 (first entry)		
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 33438.		
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 33438.		
XX	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway;		
KW	metabolic pathway; promoter; termination sequence; ss.		
XX	Arabidopsis thaliana.		
OS	Arabidopsis thaliana.		
XX	EP1033405-A2.		
PN	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
PF	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130891.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132484.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	06-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136392.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		
PR	04-JUN-1999; 99US-0137502.		
PR	07-JUN-1999; 99US-0137724.		
PR	08-JUN-1999; 99US-0138094.		

PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 32.1%; Score 422.2; DB 21; Length 1189;  
Best Local Similarity 66.8%; Pred. No. 9.9e-102;  
Matches 623; Conservative 0; Mismatches 298; Indels 12; Gaps 1;  
QY 197 CCAAGGGTGATCTCCGACAAGGCCTCGATGCGGAGCTGGTCTCGATGCGGGCCCCAG 256

Db 43 CCAGAGTAATCAGAGACAAAGACTCTATGAGAAATGGTCGAGCCATGAGATCACA 102  
QY 257 GGCAAGCTCATGGGCTGGTCCCAACCATGGCTTCCTCCACGGGGCCACCTCTCGCTC 316  
Db 103 GGCAAGACCATAGGTTTAGTACCAACAATGGTTACCTCCAGGAAGTCACTTATCTCTC 162  
QY 317 GTGGCCAGGCCCGCCAACTCTCCGACGTGCTGCGCGTCTCCATCTACGTCAACCCGGGC 376  
Db 163 GTTCGTCATCACTAGCCCTCACTGACGTCAACGTCGTCTCAATCTACGTCAATCCAGGC 222  
QY 377 CAATTCCGCCCCCAGGAGGACCTCTCCACCTACCCCTCCGACTTCGACGGCGACGTAAAG 436  
Db 223 CAATTCTCTCCCACTGAAGACCTCTCCACATACCCATCTGATTTCTCCGGCATCTCACT 282  
QY 437 AAACTCGCGTCCGTTCCCGGCGGCGTCGACGTGTTTTCCATCCCGTAACCTTGACGAT 496  
Db 283 AAGCTCGCGGCTCTTTCCGGTGGTAAAGTCGTCGTCTTTAAACCCCAAAAACCTCTATGAT 342  
QY 497 TACGG-----GAAGAACGGTGGTGACGTGGCAGAGGCTGGTGGAATGGTG 544  
Db 343 TACGGCGGTGAGACGAAGAAGATAAACGACGGTGGTGGTAATGGTGGAGGGTAGTGAGT 402  
QY 545 TCGTGGTGTGAGAGTGGTCCGGGCACGAAAGTTGGGTGAGGGTTGAGAAAGCTGGAATTG 604  
Db 403 TGTGTGGAGGAAGTGGTTAGGSCATGAGACTTGGATTAGGGTTGAGAGATTGGAGAAA 462  
QY 605 GGGCTGTGTGGGAAGACGAGGCCCGGTTTTCTTCAGAGGGTGGCGACTGTGGTGACGAAG 664  
Db 463 GGYTTGTGTGGGAAGAGTAGGCTGTGTTCTTTAGAGGTGTGCTACTATTGTTACTAAG 522  
QY 665 TTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGG 724  
Db 523 CTTTTTAATATTGTGAGCCTGATGTTGCTCTGTTGTTGTAAGAAAGATTATCAACAATGG 582  
QY 725 CGGTTTATTCAGAGGATGTTTCGAGATCTTGATTTTCCATAAAAGTGATAGTGTGAA 784  
Db 583 AGGATTATACAGAGAATGGTTCAGATCTTAATTTGGTATTGAGATTGTAGGATCAGAT 642  
QY 785 ATAAACACGTGATAATGATGCGCTGGCAATGAGTTACGTAATGTGCACCTTTACCTGAA 844  
Db 643 ATAGCTAGAGAAAAGATGGAATGCGATGAGTTTCGAGAAATGTGCGTCTTTTCAGATGAA 702  
QY 845 GAGAGGGAAGAGCACTATCAATAAATAATCAATTGTTAAGAGCTAAATCAGCAGCAGGA 904  
Db 703 GAAAGGCAAGGGCTTTGTCTATAAGTAGGTCACTGGCCATGGCTAAAGCTTCTGTGCA 762  
QY 905 GATGGTCAGGTGCAATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTGTACTGATGCT 964  
Db 763 GAAGGGAACCAATTTGCGAGAGCTTAAGGATATGATCATTCACAAAGTTGTTGGATCT 822  
QY 965 GGTGGAAGGATCGATTATGTGAGATTGTTGATCAAAATAATTTGGAGAAAAGTGGAAACAG 1024  
Db 823 GCAGGACGAGTAGATTATGTCGAGATAGTTGACCAAGAAACTCTTTGAAGGAGTAGAAGAG 882  
QY 1025 ATCAAGAGTCCGTGCTCTGTTGTTGCTGCTGATGTTTGGCAAAGTCAGGCTTATAGAC 1084  
Db 883 ATAAAGAGTGGGGTAGTGATTGTTGCTGCTGCTGTTGGAACGGTCAGGCTCATAGAC 942  
QY 1085 AACATGGAATCAACTTGTCAATGAATGTTTGA 1117  
Db 943 AACATTGAGATCAATGTCTCTCTCTAGATCTCA 975

RESULT 4  
AAZ20927  
ID AAZ20927 standard; DNA; 1264 BP.  
XX  
AC AAZ20927;  
XX  
DT 02-DEC-1999 (first entry)  
XX  
DE Nucleotide sequence of Rice Pantothenate Synthetase (PS).

XX  
KW  
KW  
XX  
OS  
XX  
FH  
FT  
FT  
FT  
XX  
PN  
XX  
PD  
XX  
PF  
XX  
PR  
PR  
XX  
PA  
XX  
PI  
XX  
DR  
DR  
XX  
PT  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
SQ

pantothenate synthetase; cloning; herbicide; biosynthesis;  
Coenzyme A precursor; pantoate-beta-alanine ligase; ds.  
Oryza sativa.

Key Location/Qualifiers  
CDS 15..1016  
/\*tag= a  
/product= Pantothenate Synthetase

WO9942565-A1.  
26-AUG-1999.  
02-JUN-1998; 98WO-EP03261.  
31-MAY-1997; 97GB-0011163.  
27-JUN-1997; 97GB-0013477.

(AGRE ) HOECHST-SCHERING AGREVO GMBH.

Abell C, Smith AG, Genschel U, Laber B;

WPI; 1999-527466/44.  
P-PSDB; AAY42403.

New isolated DNA molecule, useful in herbicidal compositions -  
Claim 7; Fig 2.2; 64pp; English.

This is the nucleotide sequence of the Oryza sativa (rice) Pantothenate synthetase (PS) enzyme, which is essential in the biosynthesis of Coenzyme A.  
The invention provides methods to assay proteins for PS activity, and the ability of compounds to inhibit PS.  
The assay compounds which are found to inhibit PS can be used as herbicides.

The nucleotide probe can be used to amplify PS coding sequences from a chosen organism via the process of polymerase chain reaction.  
A number of assays have previously been used to measure PS activity but are unsuitable for large screen biochemical screening of compounds to find PS inhibitors. In this invention however the isolation of the DNA molecule encoding PS enables a host cell to produce the protein which can then be tested against compounds potentially able to inhibit the enzyme.

Sequence 1264 BP; 279 A; 349 C; 359 G; 277 T; 0 other;

Query Match 26.9%; Score 354; DB 20; Length 1264;  
Best Local Similarity 63.2%; Pred. No. 1.3e-83;  
Matches 584; Conservative 0; Mismatches 325; Indels 15; Gaps 2;

QY 187 GGCTCCAGCCCCCAAGGTGATCTCCGACAAGGCCTCGATGCGGAGCTGGTCGCGCTCGAT 246  
Db 83 GCCGCGGAGCCGAGGTGATCCGCGACAAGGCGCGATGCGCGCATGGTCGCGCCCG 142  
QY 247 GCGGGCCAGGCAAGCTCATTTGGGTGGTTCCTCCACCATGGGTTCTCCACGGGGCCA 306  
Db 143 TCGCGCCGAGGCAAGACCGTCGCGGTCTGTACCCACCATGGGTACCTCCACCAAGGCCA 202  
QY 307 CCTCTCGCT-----CGTGGCCAGGCCCCGCCAACTCTCCGACGTCTGCGCGT 354  
Db 203 CCTCTCCCTCATCTCCGCGCGCGCGCGCGCTCCGCTGATCCCGTCGCCATCGTCT 262  
QY 355 CTCCATCTACGTCAACCCGGGCCAATTGCCCCCCCCACGGAGGACCTTCCACCTACCCCTC 414  
Db 263 CACCATCTACGTCAACCCGAGCCAGTTCCGGGCCCTCAGAGGACCTTCGCCACCTACCTTC 322  
QY 415 CGACTTCGACGGCGACGTAAAGAACTTCGCGTCCGTCGCTCCCGGGCGGTTCGACGTCTTT 474  
Db 323 CGACTTCGCGGTGACCTCCGCAAGCTTCGCTCC---ACCGGCGTCTGGATGCCGTCTT 379



QY	475	CCATCCCCGTAAC	TTGTACGATTACGG	AAGAACGGTGGTGGT	GACGTGGCAGAGG	CTGG	534
Db	380	CAACCCCTGACCT	CTACGTCCGTGGCG	CGCGGTGCGCGGG	CCGGCTCCGGAGCGC		439
QY	535	TGGAATGGTGT	CGTGCCTTGAGAGT	GGTCCGGGCACGAAAGT	TGGTGAGGTTGAGAA		594
Db	440	GATCTCTGCTG	GAGGAGGCGCGGG	ATGGCACGAGACGT	TGGGTTCCGGTGGAGCG		499
QY	595	GCTGGAATTGGG	CTGTGTGGGAAGAC	GAGCCCGCTTTCTT	CAGAGGGTGGCGACTGT		654
Db	500	ATTGAGAAGGAT	TGTGCGGGCCAGCC	GTCCCGTGTCTT	CCGAGGCGTGGCCACCAT		559
QY	655	GGTGACGAAGT	TGTTTAATATTGT	TGGAGCCAGATGGCTGT	TTCGTTAAGAGGATTA		714
Db	560	AGTCTCAAGCT	GTTTAACATCATCG	AGCCGGATGTTCCCTGT	GTTTCGGAAGAAGGATTA		619
QY	715	TCAGCAGTGGCG	GCTATTTCAGAGGAT	GGTTCGAGATCTTGAT	TTTTTCCATATAAAGTGAT		774
Db	620	TCAGCAGTGGCG	GTCACTCTGCCGTAT	TGGTCGGGACTTGAT	TTTGGCATAGAGATAAT		679
QY	775	AGTGCTGAAATA	ACACGTGATAATGAT	GGCCTGGCAATGAGTTC	ACGTAATGTGCACCT		834
Db	680	GGGATCAAGAA	ATTGTGCGAGAACTG	ATGGTCTTGCCATGAAC	TCCCGGAATGTGCACCT		739
QY	835	TTCACCTGAAG	AGAGGGGAAAGGC	ACTATCAATAATAAAT	CATTGTTAAGAGTTAAATC		894
Db	740	ATCAGCGAGGA	AGGGAAGGCATTAT	CCATCAGTAGATCACT	GGTTGATGTAGAAC		799
QY	895	AGCAGCAGGAG	ATGGTCAGGTGCATT	GTGAGAAAGTTGACAAA	ATTGGTCATCCAAAGTGT		954
Db	800	TGGCGCCTTGA	AGGGAACACTGATT	CCAAACAAATCAAA	AATAGTACAGACACT		859
QY	955	TACTGATGCTG	TGGAAGGATCGATT	ATGCTGAGATTGTGAT	CAAAATAATTTGGAGAA		1014
Db	860	AACTGAAACTG	CGGTCAGTTGACTAT	GTTTGAGATCGTGG	ACAGAAAGTTTGGTCCC		919
QY	1015	AGTGAACAGAT	CAAGAGTCCTGTCGT	CTTCTGTGTTGCTGCAT	GTGTTGGCAAAGTCAG		1074
Db	920	TGTAGAACAGAT	CGACGCCCTGTGGT	CAATTTGCGTTGCGGCG	TGGTTTGAAAGGTCAG		979
QY	1075	GCTTAGACAACA	TGGAATCAA	1098			
Db	980	GCTGATCGATA	ATAATCGAAATCGA	1003			
RESULT 5							
AAC52194							
ID	AAC52194 standard; DNA; 908 BP.						
XX	AAC52194;						
AC							
XX							
DT	18-OCT-2000 (first entry)						
XX							
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 70566.						
XX							
KW	Hybridisation assay; genetic mapping; gene expression control;						
KW	protein identification; signal transduction pathway;						
KW	metabolic pathway; promoter; termination sequence; ss.						
XX							
OS	Arabidopsis thaliana.						
XX							
PN	EP1033405-A2.						
XX							
PD	06-SEP-2000.						
XX							
PF	25-FEB-2000; 2000EP-0301439.						
XX							
PR	25-FEB-1999; 99US-0121825.						
PR	05-MAR-1999; 99US-0123180.						
PR	09-MAR-1999; 99US-0123548.						
PR	23-MAR-1999; 99US-0125788.						
PR	25-MAR-1999; 99US-0126264.						
PR	29-MAR-1999; 99US-0126785.						

PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	09-JUL-1999;	99US-0142803.
PR	12-JUL-1999;	99US-0142920.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143542.
PR	15-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.









XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
XX WPI; 2001-316495/33.  
XX P-PSDB; AAG82380.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX useful for vaccinating against infections, e.g. endocarditis -  
XX Claim 8; Page 509; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
XX (I) and (II) can have antibacterial activity and therefore can be used  
XX in vaccination. The nucleic acids (I) may be used to produce the  
XX S. epidermidis polypeptides (II) via the production of vectors  
XX containing them which are used to produce hosts cells which express the  
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
XX used to vaccinate subjects and to raise antibodies against the bacteria.  
XX The polypeptides may also be used to assay for other inhibitors of their  
XX activity and therefore identify compounds that may be used for the  
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
XX AAH5090 represent specifically claimed S. epidermidis genomic DNA  
XX polynucleotide sequences from the present invention. AAH55091 to  
XX AAH5098 represent oligonucleotide sequences and primers which are used  
XX in the exemplification of the present invention.  
XX N.B. The present invention specifically claims all the polynucleotide  
XX sequences given in the sequence listing of the present specification,  
XX however the sequence listing only goes up to SEQ ID NO:4454 so even  
XX though the sequences are given in the disclosure for SEQ ID NO:4472,  
XX no sequences are present for SEQ ID NO:4455 to 4464.  
XX Sequence 690 BP; 242 A; 95 C; 144 G; 209 T; 0 other;  
XX  
XX Query Match 7.6%; Score 99.8; DB 22; Length 690;  
XX Best Local Similarity 51.1%; Pred. No. 3.2e-16;  
XX Matches 261; Conservative 0; Mismatches 247; Indels 3; Gaps 1;  
XX  
QY 582 TGAGGGTTGAGAAGCTGGAATTGGGGCTGTGTGGGAAGAGAGGCCCGCTTTCTTCAGAG 641  
Db 152 TGAAGTTGGACACTTGGCACAAGTATTAGAGGGAGCACAAGACCTGGACACTTCGAAG 211  
QY 642 GGGTGGCGACTGTGGTGACGAAGTTGTTAATATTGTGGAGCCAGATGTGGCTGTTG 701  
Db 212 GTGTTGTGACCGTGGTCAACAACTATTAAATATTGTGCAACAGATTATGCCATTATTG 271  
QY 702 GTAAGAAGGATTATCAGCAGTGGCGGCTTATTTCAGAGGATGTTGCGAGATCTTGATTT 761  
Db 272 GGAAGAAAGGATGCAACAATATTAGCTATTGTTGAAAGAGATGTTAAAGACTTTAATCTTC 331  
QY 762 CCATAAAGTATAGGTGCTGAAATAACACGTGATAATGATGGCCTGGCAATGAGTTTCA 821  
Db 332 CTGTACATGTTATCGGTATTGATATCGTAAGAGAGAAAGATGTTAGCCAAAGCTCTA 391  
QY 822 GTAATGTGCACCTTTCACCTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 881  
Db 392 GAAATATTTACTTGACCTCTGAAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 451  
QY 882 TAAGAGCTAAATCAGCAGCAGGAGATGTTGAGTGCATTGTGAGAGATTGACAAATTTGG 941  
Db 452 GCTTAGCAAGAGATTGTATGAAGCGGGTGAACGAGATAGCAATGAGATTATAGGTCGAA 511  
QY 942 TCATCCAAAGTGT--TACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATC 998  
Db 512 TCGCTGCGTATTTAAACAAAATATTAGTGGACATATTGATGATGATTGGGATTTATAGTT 571  
QY 999 AAAATAATTTGGAGAAAGTGAACAGATCAAGAGTCCCTGTCGTTCTTCTGTTGTTGTCAT 1058  
Db 572 ATCCAAATCTTATACAAATCAAGATTTCATGGACGAATATTCATATCATTCATTGGCAGTTA 631  
QY 1059 GGTGTCGAAAGTCAGGCTTATAGACAAAT 1089

Db 632 AATTTCTAAAGCAAGATTGATAGATAATAT 662  
RESULT 10  
ABN92021  
ID ABN92021 standard; DNA; 867 BP.  
XX  
AC ABN92021;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1484.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy; gene; ds.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
(GENO-) GENOME THERAPEUTICS CORP.  
PA Doucette-Stamm LA, Bush D;  
PI  
XX WPI; 2002-381255/41.  
DR P-PSDB; ABP39476.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 1484; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 867 BP; 309 A; 123 C; 175 G; 260 T; 0 other;  
XX  
XX Query Match 7.6%; Score 99.8; DB 24; Length 867;  
XX Best Local Similarity 51.1%; Pred. No. 3.6e-16;  
XX Matches 261; Conservative 0; Mismatches 247; Indels 3; Gaps 1;  
XX  
QY 582 TGAGGGTTGAGAAGCTGGAATTGGGGCTGTGTGGGAAGAGAGGCCCGCTTTCTTCAGAG 641  
Db 329 TGAAGTTGGACACTTGGCACAAGTATTAGAGGGAGCACAAGACCTGGACACTTCGAAG 388  
QY 642 GGGTGGCGACTGTGGTGACGAAGTTGTTAATATTGTGGAGCCAGATGTGGCTGTTGTCG 701  
Db 389 GTGTTGTGACCGTGGTCAACAACTATTAAATATTGTGCAACAGATTATGCCATTATTG 448  
QY 702 GTAAGAAGGATTATCAGCAGTGGCGGCTTATTTCAGAGGATGTTGCGAGATCTTGATTT 761  
Db 449 GGAAGAAAGGATGCAACAATATTAGCTATTGTTGAAAGAGTGGTTAAAGACTTTAATCTTC 508  
QY 762 CCATAAAGTATAGGTGCTGTAATAACACGTGATAATGATGGCCTGGCAATGAGTTTCA 821  
Db 509 CTGTACATGTTATCGGTATTGATATCGTAAGAGAAAGATGTTTAGCCAAAGCTCTA 568





```
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3721..3780
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
PN EP786519-A2.
XX
PD 30-JUL-1997.
XX
PF 07-JAN-1997; 97EP-0100117.
XX
PR 05-JAN-1996; 96US-0009861.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
DR WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
PS Claim 1; Page 849-852; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
SQ Sequence 4549 BP; 1387 A; 858 C; 698 G; 1421 T; 185 other;
Query Match 7.5%; Score 99; DB 18; Length 4549;
Best Local Similarity 51.5%; Pred. No. 1.3e-15;
Matches 250; Conservative 1; Mismatches 231; Indels 3; Gaps 1;
614 GGAAGAGCAGGCCCGTTTCTTCAGAGGGTGGCGACTGTGTGACGAAGTTGTTTAAAT 673
3351 GGAGCGAACGGYCCAGGCGATTCTTGTATGGGTGGTAACAGTAGTTAATAAGCTATTAAAT 3292
674 ATTGTGGAGCCAGATGTGGCTGTGTTTCGTAAGAAGGATTATCAGCAGTGGCGGCTTATT 733
3291 ATTGTTATGCCGGAATTATGCTTATTTTGGTAAAAAAGATGCCAGCAATTGGCTATTGTC 3232
734 CAGAGGATGGTTCGAGATCTTGATTTTTCATATAAAGTAGTGTGTAATAACACGT 793
3231 GAGCAAAKGGTAAAGACTTCAATCATCGCTGGAATATTGGTATTGATATCGTTTGA 3172
794 GATAATGATGGCCTGCAATGAGTTCCAGTATGTGCACCTTTTCACTGAAGAGAGGGA 853
3171 GAAGCAGATGGTTTGGCGAAAAGTTCAAGAAATGTTTATTAAACGGAGCAAGAACGACAA 3112
```

```
QY 854 AAGGCATATCAATAAATAAATCATTTGTTAAGAGCTAAATCAGCAGCAGGATGGTCAG 913
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3111 GAAGCGGTACATTTAAGTAAAGTTTGTCTATTAGCACAAGCGTTGTATCAGACGGTGAA 3052
QY 914 GTGCATTGTGAGAAAGTTGACAAATTTGGTCAATCCAAAGTGT--ACTGATGCTGGTGA 970
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3051 CGTCAAGTAAGTAATATAGACAGGGTTACTGAATATCTTGAATCACATATAAGTGAA 2992
QY 971 AGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTGGACAGATCAAG 1030
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2991 CGCATGAAGAAGTTGCTGTTATAGTTATCTCTCAATTAGTAGAACAACACGAAATAACT 2932
QY 1031 AGTCCTGTGCTCTTCTGTGTTGCTGCTGATGGTTTGGCAAAGTCAGGCTTATACACATG 1090
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2931 GGACGCATATTATCTCGTTAGCTGTTAAATTTTCAAAAGCGCGTTTAATAGATAATATA 2872
QY 1091 GAAAT 1095
Db |||
2871 ATAAT 2867
RESULT 13
AAA81463/c
ID AAA81463 standard; DNA; 78845 BP.
XX
AC AAA81463;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_11 SEQ ID NO:11.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 330-353; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
```



Db 291037 GAATCCTTGTGTCAGGCGAGTTTGGATTATGTCAGGTTTGGAAACGTCGCGTCCAATCC 291096  
QY 953 GTTACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTGATCAAAATAATTGGAG 1012  
Db 291097 CTGACAGATACGCGTGGTGGTGCATGATGTCGAAATCCGCGCGCGATACGCTCGAA 291156  
QY 1013 AAAGTGAACAGATCAAGAGTCCTGTCGTTCTTGTGTTGCTGATGTTGGCAAAGTC 1072  
Db 291157 GTGGCGCGGGGAGAGATAAGAACTGGTGGTCTTGGCGCGCGCTGTCTGGGACGACG 291216  
QY 1073 AGGCTTATAGACAACATGGAATCAA 1098  
Db 291217 CGCCTGATTGACAATTGGAATAAA 291242

RESULT 15  
AAA81490  
ID AAA81490 standard; DNA; 1437668 BP.  
XX  
AC AAA81490;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200022430-A2.  
XX  
PD 20-APR-2000.  
XX  
PF 08-OCT-1999; 99WO-US23573.  
XX  
PR 09-OCT-1998; 98US-0103794.  
PR 30-APR-1999; 99US-0132068.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizza M;  
XX  
DR WPI; 2000-318079/27.  
XX  
PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
PT used in the diagnosis and treatment of N. meningitidis infection and  
PT other Neisserial infections, for example, N.gonorrhoea -  
XX  
PS Claim 7; Page 866-1272; 1760pp; English.  
XX

CC The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC Neisserial bacteria. For example, some of the identified proteins could  
CC be components of vaccines against Meningococcus B; against all serotypes;  
CC and/or against all pathogenic Neisseriae. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.  
XX  
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;  
Query Match 7.0%; Score 91.6; DB 21; Length 1437668;  
Best Local Similarity 48.8%; Pred.No. 1.9e-12;  
Matches 247; Conservative 0; Mismatches 259; Indels 0; Gaps 0;  
QY 593 AAGCTGGAATTGGGGCTGTGTGGGAAGAGAGAGGCGGCTTTCTTCTCAGAGGGGTGGCGACT 652  
Db 890736 AATCTGCAAAATGAGTTGTGCGGCAAAATCCGCGGCAATTTTCGCGGTGTGGCAACG 890795  
QY 653 GTGGTGACGAAGTTGTTTAATATTGTGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGAT 712  
Db 890796 GTTGTCTTCTAAATGTTCCACATCGTTTCCCGGACATTCGCTGTTGGTAAGAAGGAT 890855  
QY 713 TATCAGCAGTGGCGGCTTATTTCAGAGAGATGTTTCGAGATCTTGAATTTTCCATAAAAGTG 772  
Db 890856 TACCAGCAGCTTCCGCTGATTAAAGGTTTGTGCGAAGATTTGAATTTTGATGTTGAATA 890915  
QY 773 ATAGGTGCTGAAATAACACGTCGATATGATGCTGCGCTGGCAATGAGTTCACGTAATGTGCAC 832  
Db 890916 GTGCCTGTTGATACAGGGCGCGGGAAGACGGGTTGGCACTGTGAGCCGCAACCATAT 890975  
QY 833 CTTCACCTGAAGAGAGAGGAAAGGCACTATCAATAAATAATCATTTGTAAGAGCTAAA 892  
Db 890976 TTGAGTGGCGGACGCGACGAGCAGCCGCGCTGTACCCGCAATTAAGGCTGTTGCC 891035  
QY 893 TCAGCAGCAGGAGATGTCAGGTGCTTGTGAGAGATTGACAAATTTGGTCAATCCAAAGT 952  
Db 891036 GAATCCTTGGTGCAGGCGAGTTTGGATTATGTCAGGTTTGGAAACGTCGCGTCCAATCC 891095  
QY 953 GTTACTGATGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAG 1012  
Db 891096 CTGACAGAAATACGCTGGGTGGTTCGATTATGTCGAAATCCGCGCGCGATACGCTCGAA 891155  
QY 1013 AAAGTGAACAGATCAAGAGTCCTGTCGTTCTTGTGTTGCTGATGTTGGCAAAGTC 1072  
Db 891156 GTGGCGCGGGGAGAGATAAGAACTGTTGGTCTTGGCGCGCGCTGTCTGGGACGACG 891215  
QY 1073 AGGCTTATAGACAACATGGAATCAA 1098  
Db 891216 CGCCTGATTGACAATTGGAATAAA 891241

Search completed: December 15, 2003, 21:51:45  
Job time : 432 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: December 15, 2003, 23:15:54 ; Search time 471 Seconds  
(without alignments)  
9286.320 Million cell updates/sec  
Title: US-10-033-269-8  
Perfect score: 1316  
Sequence: 1 tcggcacgaggctttctcag.....tggcattgctatttgctagg 1316  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2201672 seqs, 1661799599 residues  
Total number of hits satisfying chosen parameters: 4403344  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1316	100.0	1316	13	US-10-033-269-8
2	506	38.4	1183	13	US-10-033-269-6
3	389.8	29.6	1831	13	US-10-033-269-3
4	384.4	29.2	1235	13	US-10-033-269-12
5	382.6	29.1	1443	13	US-10-033-269-1
6	346.2	26.3	1148	13	US-10-033-269-10
7	300.2	22.8	1644	13	US-10-033-269-5
8	118.8	9.0	9381	15	US-10-114-170-7
9	109	8.3	1672	10	US-09-070-927A-757
10	99	7.5	4549	8	US-08-781-986A-174
11	83.6	6.4	546	10	US-09-974-300-1035
12	73.4	5.6	640681	10	US-09-790-988-1
13	70.8	5.4	837	10	US-09-738-626-132
14	70.8	5.4	3309400	10	US-09-738-626-1
15	67.2	5.1	501	10	US-09-974-300-5461

16	67.2	5.1	501	10	US-09-974-300-5490	Sequence 5490, Ap
17	57.8	4.4	999	15	US-10-156-761-4671	Sequence 4671, Ap
18	57.8	4.4	9025608	15	US-10-156-761-1	Sequence 1, Appli
c 19	52.8	4.0	520	15	US-10-184-644-332	Sequence 332, App
c 20	52.8	4.0	520	15	US-10-184-634-332	Sequence 332, App
21	52.6	4.0	1167	15	US-10-156-761-2743	Sequence 2743, Ap
c 22	52.6	4.0	9025608	15	US-10-156-761-1	Sequence 1, Appli
c 23	51	3.9	390	9	US-09-790-399-7	Sequence 7, Appli
c 24	49.4	3.8	440	13	US-10-063-685-52	Sequence 52, Appl
c 25	49.4	3.8	440	15	US-10-184-644-202	Sequence 202, App
c 26	49.4	3.8	440	15	US-10-184-634-202	Sequence 202, App
27	49.2	3.7	1392	15	US-10-156-761-3990	Sequence 3990, Ap
28	49	3.7	390	9	US-09-790-399-7	Sequence 7, Appli
29	48.8	3.7	825	15	US-10-156-761-2303	Sequence 2303, Ap
30	48.8	3.7	2274	15	US-10-156-761-5313	Sequence 5313, Ap
31	48.4	3.7	543	15	US-10-156-761-7012	Sequence 7012, Ap
32	48.2	3.7	1416	15	US-10-156-761-6191	Sequence 6191, Ap
33	48	3.6	1209	15	US-10-156-761-6772	Sequence 6772, Ap
34	48	3.6	1479	15	US-10-156-761-4461	Sequence 4461, Ap
35	47.8	3.6	4863	9	US-09-815-242-4071	Sequence 4071, Ap
c 36	47	3.6	1475	13	US-10-271-889-12	Sequence 12, Appl
37	46.8	3.6	939	15	US-10-156-761-7243	Sequence 7243, Ap
38	46.8	3.6	1428	13	US-10-289-757-149	Sequence 149, App
39	46.8	3.6	1429	13	US-10-289-757-40	Sequence 40, Appl
c 40	46.8	3.6	1440	13	US-10-289-757-43	Sequence 43, Appl
41	46.8	3.6	1749	15	US-10-100-679-58	Sequence 58, Appl
42	46.6	3.5	11058	15	US-10-156-761-3629	Sequence 3629, Ap
43	46.4	3.5	4674	13	US-09-940-316B-26	Sequence 26, Appl
44	46.4	3.5	4737	13	US-09-940-316B-30	Sequence 30, Appl
45	46.2	3.5	1440	15	US-10-156-761-6190	Sequence 6190, Ap

ALIGNMENTS

RESULT 1  
US-10-033-269-8  
; Sequence 8, Application US/10033269  
; Publication No. US20030167503A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutteridge, Steve  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Orozco, Buddy  
; TITLE OF INVENTION: Plant Genes Encoding Pantothenate Synthetase  
; FILE REFERENCE: BBI446 US NA  
; CURRENT APPLICATION NUMBER: US/10/033,269  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/247,938  
; PRIOR FILING DATE: 2000-11-13  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 1316  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-033-269-8

Query Match	100.0%;	Score 1316;	DB 13;	Length 1316;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1316;	Conservative	0;	Mismatches	0; Gaps 0;
QY	1	TCGGCAGGAGGCTTTCTCAGCAGGAGGCGCTCGTGCCTACACCAACCGTACTTCAAC	60	
Db	1	TCGGCAGGAGGCTTTCTCAGCAGGAGGCGCTCGTGCCTACACCAACCGTACTTCAAC	60	
QY	61	CCAAAGAGTCATAATAACTTGCCTAACACCTACCTCTTCAACCCGCGTTCGGCCAT	120	
Db	61	CCAAAGAGTCATAATAACTTGCCTAACACCTACCTCTTCAACCCGCGTTCGGCCAT	120	
QY	121	CACGTCGTCCTTTTCACCTTATTACACTCCTCCTACGAACTCCCACTGTTTCCAATATTATT	180	
Db	121	CACGTCGTCCTTTTCACCTTATTACACTCCTCCTACGAACTCCCACTGTTTCCAATATTATT	180	

QY 181 CACAATGGCTCCAGCCCAAGGGTGATCTCCGACAAGGCCTCGATGCGGAGCTGGTCGGC 240  
Db 181 CACAATGGCTCCAGCCCAAGGGTGATCTCCGACAAGGCCTCGATGCGGAGCTGGTCGGC 240  
QY 241 CTCGATCGGGCCCGCAGGCAAGCTCATTTGGGTGGTCCCAACCATGGGCTCCTCCACGC 300  
Db 241 CTCGATCGGGCCCGCAGGCAAGCTCATTTGGGTGGTCCCAACCATGGGCTCCTCCACGC 300  
QY 301 GGGCCACCTCTCGCTCGTGGCCCAAGCTCATTTGGGTGGTCCCAACCATGGGCTCCTCCACGC 360  
Db 301 GGGCCACCTCTCGCTCGTGGCCCAAGCTCATTTGGGTGGTCCCAACCATGGGCTCCTCCACGC 360  
QY 361 CTACGTCAACCGGGCCAATTGCGCCCAAGGAGGACCTCTCCACCTACCCCTCCGACTT 420  
Db 361 CTACGTCAACCGGGCCAATTGCGCCCAAGGAGGACCTCTCCACCTACCCCTCCGACTT 420  
QY 421 CGACGGCGACGTAAGAACTCGCGTCCGTTCCCGGGCGGTCGACGTCGTTTCCATCC 480  
Db 421 CGACGGCGACGTAAGAACTCGCGTCCGTTCCCGGGCGGTCGACGTCGTTTCCATCC 480  
QY 481 CCGTAACCTGTACGATTACGGGAAGAACGCTGGTGGTGACGTTGGCAGAGGCTGGTGAAT 540  
Db 481 CCGTAACCTGTACGATTACGGGAAGAACGCTGGTGGTGACGTTGGCAGAGGCTGGTGAAT 540  
QY 541 GGTGTCGTGCTGAGAGTGGTCCGGGACGAAAGTTGGGTGAGGTTGAGAAAGCTGGA 600  
Db 541 GGTGTCGTGCTGAGAGTGGTCCGGGACGAAAGTTGGGTGAGGTTGAGAAAGCTGGA 600  
QY 601 ATTGGGGCTGTGGGAAGAGCAGGCCCGTTTCTTCAGAGGGTGGCGACTGTGGTGAC 660  
Db 601 ATTGGGGCTGTGGGAAGAGCAGGCCCGTTTCTTCAGAGGGTGGCGACTGTGGTGAC 660  
QY 661 GAAGTTGTTAATATTGTGGAGCCAGATGGTGTTCGTTAAGAAAGGATTATCAGCA 720  
Db 661 GAAGTTGTTAATATTGTGGAGCCAGATGGTGTTCGTTAAGAAAGGATTATCAGCA 720  
QY 721 GTGGCGGCTATTTCAGAGGATGGTTCGAGATCTTGATTTTTCATAAAAGTATAGGTGC 780  
Db 721 GTGGCGGCTATTTCAGAGGATGGTTCGAGATCTTGATTTTTCATAAAAGTATAGGTGC 780  
QY 781 TGAATAACACGTCGATTAATGATGGCCCTGGCAATGAGTTACGTAATGTGCACCTTTACC 840  
Db 781 TGAATAACACGTCGATTAATGATGGCCCTGGCAATGAGTTACGTAATGTGCACCTTTACC 840  
QY 841 TGAAGAGAGGAAAAGGCACCTATCAATAAATAATCATTTGTTAAGAGCTAAATCAGCAGC 900  
Db 841 TGAAGAGAGGAAAAGGCACCTATCAATAAATAATCATTTGTTAAGAGCTAAATCAGCAGC 900  
QY 901 AGGAGATGGTCAGGTGCAATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTTACTGA 960  
Db 901 AGGAGATGGTCAGGTGCAATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTTACTGA 960  
QY 961 TGCTGTGGAAGGATCGATTATGCTGAGATTTGTTGATCAAAATAAATTTGGAAAGTGA 1020  
Db 961 TGCTGTGGAAGGATCGATTATGCTGAGATTTGTTGATCAAAATAAATTTGGAAAGTGA 1020  
QY 1021 ACAGATCAAGAGTCCGTGCTCTTCTGTTGCTGCAATGGTTGGCAAAGTCAGGCTTAT 1080  
Db 1021 ACAGATCAAGAGTCCGTGCTCTTCTGTTGCTGCAATGGTTGGCAAAGTCAGGCTTAT 1080  
QY 1081 AGACBACATGGAATCAACTTGTCAATGAATGTTTGTGATCTAACCTTCTGTCACTCAAAAC 1140  
Db 1081 AGACBACATGGAATCAACTTGTCAATGAATGTTTGTGATCTAACCTTCTGTCACTCAAAAC 1140  
QY 1141 ATGGCCACATGCTTAATAATAGTTCCGGCCACGTCGTTAACAATTTCTAACAGTTTCATG 1200  
Db 1141 ATGGCCACATGCTTAATAATAGTTCCGGCCACGTCGTTAACAATTTCTAACAGTTTCATG 1200  
QY 1201 GTTATAGTCATGACAAATTTTCTGTCAGCCATACATGATTACTTGTAGATGCAATTT 1260  
Db 1201 GTTATAGTCATGACAAATTTTCTGTCAGCCATACATGATTACTTGTAGATGCAATTT 1260  
QY 1261 TACCGCATCAAAATTTCTATGAGAGCTGCTACCAAGTGGCATTTGCTATTTTGCTAGG 1316

Db 1261 TACCGCATCAAAATTTCTATGAGAGCTGCTACCAAGTGGCATTTGCTATTTGCTAGG 1316  
RESULT 2  
US-10-033-269-6  
; Sequence 6, Application US/10033269  
; Publication No. US20030167503A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutteridge, Steve  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Orozco, Buddy  
; TITLE OF INVENTION: Plant Genes Encoding Pantothenate Synthetase  
; FILE REFERENCE: BB1446 US NA  
; CURRENT APPLICATION NUMBER: US/10/033,269  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/247,938  
; PRIOR FILING DATE: 2000-11-13  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 6  
; LENGTH: 1183  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-10-033-269-6  
Query Match 38.4%; Score 506; DB 13; Length 1183;  
Best Local Similarity 72.1%; Pred. No. 6.4e-142;  
Matches 691; Conservative 0; Mismatches 255; Indels 12; Gaps 2;  
QY 197 CCAAGGGTGATCTCCGACAAGGCCTCGATGCGGAGCTGGTCGCGTCCGATGCGGGCCAG 256  
Db 144 CCCTGGTCATCACCGACAAGCCGAGATCGGGAGTGGTCCGAGTCCGCGGGCCAG 203  
QY 257 GGCAAGCTCATTTGGGTGGTCCCAACCATGGGCTTCTCCACGCGGGCCACCTCTCGTC 316  
Db 204 GGCAAGCTCATTTGGGTGGTCCCAACCATGGGCTTCTCCACGAGGGCCACCTCTCGTC 263  
QY 317 GTGGCCAGGCCCGCAACTCTCCGACGTCGTCGCGCTCTCCATCTACGTCACCCGGGC 376  
Db 264 GTCCGGAGGCCCGCGCGCGCGCGCGCGCTCGTGGTCTCCGTCACGTCGACCCCGGC 323  
QY 377 CAATTGCCCCCAGGAGGACCTCTCCACCTACCCCTCCGACTTCGACGCGGCGAGTAAAG 436  
Db 324 CAGTTCGCCCCCTCCGAGGACCTCTCGACGTCACCATCTGATTTTCGAGGGCGACCTGGGC 383  
QY 437 AAATTCGCGTCCGTTCCCGCGCGCGCTCGACGTCGTTTCCATCCCGTAACTGTACGAT 496  
Db 384 AAGCTCAGGGCCGTCGCGCGCGCGCTGGACGTCGTTTCTCGTCCCAAGATCTTACGAC 443  
QY 497 TACGGGAAGAACGGTGGTGGTGAC-----GTGGCAGAGGCTGGTGAATGGTGTG 547  
Db 444 TACGGTCAACGCGAGGTCGGTGGCTCTGGCGTTCGAGCGGATAATGGTCCGTCCTGT 503  
QY 548 TGGCTTGAGAGTGGTCCGGGCGACGAAAGTTGGGTGAGGTTGAGAAAGCTGGAATTTGGG 607  
Db 504 TTGAGGAGAAAGGCGATGGGCGACGAGCGGTGGGTGAGAGGTTGGAGAGGCGG 563  
QY 608 CTGTGTGGGAAGAGCAGGCCCGCTTTCTTCAGAGGGGTGGCGACTGTGTCGACGAAGTTG 667  
Db 564 ATGTGCGGGAAGAGCAGGCCCGCTGTTCTTTTCGAGGGGTGGCCACTGTGGTCACCAAGCTG 623  
QY 668 TTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAGGATTATCAGCAGTGGCGG 727  
Db 624 TTCAACATTTGTGAGCCGGATGTGAGTGTTCGGAAGAAAGATTACCAAGTGGCGG 683  
QY 728 CTTATTTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAGTATAGGTGCTGAAATA 787  
Db 684 ATCATTCGGCGGTGGT---GAATCTTGACTTTCCATAACAAGTATAGGTCTTGAGGTC 740  
QY 788 ACAGTGTATAATGATGGCCCTGGCAATGAGTTACGTAATGTGCACCTTTACCTGAAGAG 847  
Db 741 ATCGAGATCATGATGGCCCTTGGCTAAGCTCAGCAATGTGCAATGTGCACTCTCACCTGAAGAA 800









Db 504 TTGAGGAGAGGGCATGGGACAGGGCGTGGGTGAGGGTGGAGAGGTTGGAGAAGGGC 563  
QY 608 CTGTGTGGGAAGAGCAGGCCCGTTTCTTCAGAGGGGTGGCGACTGTGGTGACGAAGTTG 667  
Db 564 ATGTGCGGAAGAGCAGGCCCGTGTCTTTCGAGGGGTGGCCACTGTGGTCAACCAAGCTG 623  
QY 668 TTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTTAAGAAGGATTATCAGCAGTGGCGG 727  
Db 624 TTCAACATTGTGGAGCCGATGTGAGTGTTTTCGGGAAGAAGGATTACCAGCAGTGGCGG 683  
QY 728 CTTATTCAGAGGATGGT 744  
Db 684 ATCATTCGGCGGTTGGT 700

RESULT 8  
US-10-114-170-7/c  
; Sequence 7, Application US/10114170  
; Publication No. US20030023075A1  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; Burland, Valerie  
; Perna, Nicole T.  
; Plunkett, Guy  
; Welch, Rod  
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; City: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/114,170  
; FILING DATE: 01-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/453,702  
; FILING DATE: 03-DEC-1999  
; APPLICATION NUMBER: 60/110,955  
; FILING DATE: 04-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 251-5000  
; TELEFAX: (608) 251-9166  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9381  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-10-114-170-7

Query Match 9.0%; Score 118.8; DB 15; Length 9381;  
Best Local Similarity 48.3%; Pred. No. 4.1e-24;  
Matches 427; Conservative 0; Mismatches 397; Indels 60; Gaps 1;  
QY 239 CGCTCGATCGGGCCAGGGCAAGCTCATTGGGTGGTCCCCACCATGGGCTTCCTCCAC 298  
Db 1111 CGCCGCCTGCGTATGGAAGGCAAGCGGTGGCGCTGGTACCTACCATGGGCAACCTGCAT 1052

QY 299 GCGGGCCACCTCTCGCTCGTGGCCCCAGGCCCGCAAACTCTCCGACGTGCTCGCCGCTCTCC 358  
Db 1051 GATGGCCACATGAAGCTGGTCGACGAAGCCAAAGCCCGCGCATGTGCTCGTCACT 992  
QY 359 ATCTACGTCAACCCGGGCCAAATTCGCCCCCAAGGAGGACCTCTCCACCTACCCCTCCGAC 418  
Db 991 ATTTTCGTTAACCCGATGCAGTTCGACCCCGGAGGACCTGGCTCGGTACCCACGACC 932  
QY 419 TTCGACGGCGACGTAAAGAAACTCGCGTCCGTTCCCGCGCGTCGACGTCGTTTCCAT 478  
Db 931 TTGCAGGAAGACTCGGAAAAGCT----- 909  
QY 479 CCCCCTAACTTGACGATTACGGGAAGAACGGTGGTGACGTGGCAGAGGCTGGTGA 538  
Db 908 -----GAACAAGCGCAAGTGGA TCTGGTTTCGCGCCCTCG 872  
QY 539 ATGGTGTGCTGCGTTGAGAGTGGGTCCGGGCACGAAAGTTGGGTGAGGTTGAGAAAGCTG 598  
Db 871 GTAAAGAGATCTACCCGAACGGTACTGAAACCCATACTTACGTTGACGTTCCAGGCTT 812  
QY 599 GAATTGGGGCTGTGTGGGAAGAGCAGGCCCGCTTTTCTTCAGAGGGGTGGCGACTGTGGTG 658  
Db 811 TCGACCATGCTGGAAGGCGCCAGCCGCGGGGCATTTCGCGCGGTGTCGACTATCGTC 752  
QY 659 ACGAAGTTGTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTTAAGAAGGATTATCAG 718  
Db 751 AGCAAGCTGTTAACTGTTCCAGCCCGGACATCGCTGCTTCGGTGAAAAGGACTTTCAG 692  
QY 719 CAGTGGCGGCTTATTCAGAGGATGGTTCGAGATCTTGATTTTTCCATAAAAGTGATAGT 778  
Db 691 CAACTGGCGCTGATCCGCAAAATGGTTGCTGATATGGGCTTTGATATTGAGATTGTCGT 632  
QY 779 GCTGAAATAACACGTGATAATGATGGCTGGCAATGAGTTACGTAATGTGCACCTTTCA 838  
Db 631 GTGCCAATTATGCGCGCCAAAGACGGTCTGSCACTGAGTTCCCGTAACGGTTATCTGACG 572  
QY 839 CCTGAAGAGAGGGAAAAGGCACATATCAATAAAATAAATCATTTGTTAAGAGCTAAATCAGCA 898  
Db 571 GCAGAACAACGCAAAATTCGCCCGGCTGTACAAAGTTTAAAGTCGATTGCCGACAAA 512  
QY 899 GCAGGAGATGTCAGGTGCATTGTGAGAAAGTTGACAAATTTGGTCATCCAAAGTGTACT 958  
Db 511 TTGCAGGCTGGCGAACGGGATCTCGATGAAATATTGCTATTGCGCGGCAAGAACTGAAT 452  
QY 959 GATGCTGTGGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTG 1018  
Db 451 GAAAAGGCTTCGCGCGCGATGATATTCAGATTTCGGATGCCGACACATTGCTGGAAGTC 392  
QY 1019 GAACAGATCAAGAGTCCTGTCGTTCTTCTGTGTTGTCATGTTGGCAAGTCAGGCTT 1078  
Db 391 TCTGAAACCAGCAAAACGGGCAGTAATTCTGGTAGCCGCTGGCTGGCGATGCTCGCCTG 332  
QY 1079 ATAGACAACATGGAATCAACTTGTCAATGAATGTTTGTATCTAA 1122  
Db 331 ATCGACAACAATAATGGTCGAGCTGGCGTAATACTTATACTCTAA 288

RESULT 9  
US-09-070-927A-757/c  
; Sequence 757, Application US/09070927A  
; Patent No. US20020120116A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles A. Kunsch  
; Patrick J. Dillon  
; Steven Barash  
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 982  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA



ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 757:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1672 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 757:  
US-09-070-927A-757

Query Match 8.3%; Score 109; DB 10; Length 1672;  
Best Local Similarity 51.1%; Pred. No. 1.1e-21;  
Matches 275; Conservative 2; Mismatches 258; Indels 3; Gaps 1;  
QY 605 GGGCTGTGTGGGAAGAGCAGGCCGCTTTCTTCAGAGGGGTGGCGACTGTGTGACGAAG 664  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :  
593 GGTATGTGGAGCTAGTAGGCCAATACATTTTAGAGGGGTTGCACTGTGTGAAGTAAR 534  
QY 665 TTGTTTAATATGTGGAGCCAGATGTGGCTGTGTTGCGTAAGGAAGGATTATACAGCAGTGG 724  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
533 TTATTTAATACATAATACCAGCAGATAGAGCTTATTTTGGAGAAAGGATGCACACAGTTA 474  
QY 725 CGGCTTATTCAGAGGATGTTTCGAGATCTTGATTTTCCATAAAAGTGATAGGTGCTGAA 784  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
473 GCTGTAATTAAGAAATGTTTAGGACCTTAATATTGATATAGACGTTGTAGGGTGTCCA 414  
QY 785 ATAACACGTGATAATGATGGCCCTGCCAATGAGTTTACGTAATGTGCACCTTTTCACCTGAA 844  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
413 ATTATAAGGAAGAGGATGGACTAGCTAAAGTTCTAGAAACACATACCTTAAGTTTAGAG 354  
QY 845 GAGAGGGAAGGCACTATCAATAAATCAATCTTGTAAAGAGCTAAATCAGCAGCAGGA 904  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
353 GAAAGATCTCAGCTACTATTTTAAATAAATCTTTAACTTTAGCYAAGGAAGCCTTAAAT 294  
QY 905 GATGTCAGGTGCATTGTGAGAGATTGACAAATTTGGTCATCCAAAGTGT---TACTGAT 961  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
293 AATGGAGAGAGATAGCTYAAATAATATKAGATAAATAGTAAATAATAAATACTTGT 234  
QY 962 GCTGTGGAAGATCGATTATGCTGAGATTGTTGATCAAAATAAATTTGGAGAAAGTGGAA 1021  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
233 AATTTGGCAAGATTGATTATGTAGAGTTGTTGATTTCTTTATCATTACAAAGAGTAAAT 174  
QY 1022 CAGATCAAGAGTCCTGCTCTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
173 TATATTGAAAATCAGTTTGTAGTTGCAATAGCAGTATTTTATAGGAAACACAGATTAAATA 114  
QY 1082 GACAACATGGAATCAACTTGTCAATGAATGTTTGTATCTAACCTTCTGTCTATCTCAAA 1139  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
113 GATAATTTTACATTTGAATTACAATAGGGGGCTATAATAATGANAGTTCATGTTAAA 56

RESULT 10  
US-08-781-986A-174/c  
Sequence 174, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 174:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4549 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-174

Query Match 7.5%; Score 99; DB 8; Length 4549;  
Best Local Similarity 51.5%; Pred. No. 2.4e-18;  
Matches 250; Conservative 1; Mismatches 231; Indels 3; Gaps 1;  
QY 614 GGAAGAGCAGGCCCGCTTTCTTCAGAGGGGTGGCGACTGTGTGACGAAGTTGTTTAAAT 673  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3351 GGAGCGAACGGYCCAGGGCATTTTGATGGGGTGGTAACAGTAGTTAATAAGCTATTAAAT 3292  
QY 674 ATTGTGGAGCCAGATGTGGCTGTGTTTCGGTAAGAGGATTATCAGCAGTGGCGGCTATT 733  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3291 ATTGTATGCCGATTATGCTTATTTTGGTAAAGATGCCAGCAATTGGCTATTGTC 3232  
QY 734 CAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAGTGATAGGTGCTGAATAACACGT 793  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3231 GAGCAAKGGTAAAGACITCAATCATGCCGTTGAAATATTGGTATTGATATCGTTCCA 3172  
QY 794 GATAATGATGCCCTGGCAATGAGTTCAGTAAATGTGACCTTTCACCTGAAGAGAGGAA 853  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3171 GAAGCAGATGGTTTGGCGAAAGTTCAAGAAATGTTTATTAAACGGAGCAAGAACGACAA 3112  
QY 854 AAGGCACTATCAATAAATAATCATTTGTTAAGAGCTAAATCAGCAGCAGGAGATGGTCAG 913  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3111 GAAGCGGTACATTTAAGTAAAGTTGCTATTAGCAAGAGCTTGTATCAAGAGCGGTGAA 3052  
QY 914 GTGCATTGTGAGAGTTGACAAATTTGGTCATCCAAAGTGT---ACTGATGCTGTGGA 970  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
3051 CGTCAAACTAAAGTAATTATAGACAGGGTTACTGAATCTTGAATCACATATAAGTGAA 2992  
QY 971 AGGATCGATTATGCTGAGATTGTTGATCAAAATAATTGGAGAAAGTGGAACACATCAAG 1030  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2991 CGCATTGAAGAAGTTCGTGTTTATAGTTATCTCAATTAGTAGAACAACACGAAATAACT 2932  
QY 1031 AGTCCTGTCGTCCTTCTGTTGTTGTCATGTTTGGCAAAGTCAGGCTTATAGACAAACATG 1090  
Db 2931 GGACGCATATTATCTCGTTAGCTGTTAAATTTCAAAGCGCGTTTAATAGATAATATA 2872  
QY 1091 GAAAT 1095  
Db 2871 ATAAT 2867

RESULT 11  
US-09-974-300-1035  
; Sequence 1035, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1035  
; LENGTH: 546  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-1035

Query Match 6.4%; Score 83.6; DB 10; Length 546;  
Best Local Similarity 47.8%; Pred. No. 2.5e-14;  
Matches 242; Conservative 0; Mismatches 264; Indels 0; Gaps 0;  
QY 584 AGGTTGAGAAGCTGGAATGGGGCTGTGGGAGAGCAGGCCCGTTTCTTCAGAGGG 643  
Db 27 AGCGTGAAGAAAGCGGACGGATGTCCTTTGCGCGCTTCAAGAGAAGGCAATTCGACGGT 86  
QY 644 GTGGCACTGTGGTGACGAAGTGTGTTAATATTGTGGAGCCAGATGTGGCTGTGTTGGT 703  
Db 87 GTCGCAACAGTGCTCACAAGCTGTTCATCTCACGTCTCCGACAGGGTTTATTCGGC 146  
QY 704 AAGAAGGATTATCAGCAGTGGCGCTTATTTCAGAGGATGGTTCGAGATCTTGATTTTCC 763  
Db 147 ATGAAGGATGCCAGCAGGTTCGGCTCGTCGACGGGCTGATCAACGACTTTTTCATGGAT 206  
QY 764 ATAAAGTGATAGGTGCTGAATAACACGTGATAATATGCGCTGGCAATGAGTTCACGT 823  
Db 207 ATCGAGCTGATCCCTGTGGAACAACAAACGGGAAGACGGGCTTGCCAAAAGCTCAAGA 266  
QY 824 AATGTGCACCTTTCACCTGAAGAGAGGGAAAGGCACCTATCAATAAATAATCATTTGTTA 883  
Db 267 AATGTCAATCTCGAGCTGAAGAACGGCAGGAGACCGGCGCTTTACCGAGCTCTTCAG 326  
QY 884 AGAGCTAAATCAGCAGCAGGAGATGGTCAGGTGCAATTGTGAGAAGTTGACAAATTTGGTC 943  
Db 327 CGCGGCGCGAGCTGATCCGCAATGGAGAGAGACCGGAGCGGCTTAAGCGGAGATT 386  
QY 944 ATCCAAAGTGTACTGATGCTGTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAAT 1003  
Db 387 CGGAGCATCTGGAAAAAACGAGTGGTGTATTGATTATGCCGATATTTATTCATATCCC 446  
QY 1004 AATTGGAGAAAGTGGAAACAGATCAAGAGTCCCTGTCCTCTTCTGTGTTGCTGATGGTTT 1063  
Db 447 GATCTGGAGATCAGGGATCCATTGACTGGAAAAAGTCATCAGCGGTTGCCGTTCAAGTTT 506  
QY 1064 GGCAAAGTCAGGCTTATAGACAACAT 1089

Db 507 TCCAAAGCCCGCTTAATTGACAATAT 532  
RESULT 12  
US-09-790-988-1/c  
; Sequence 1, Application US/09790988  
; Patent No. US20020127687A1  
; GENERAL INFORMATION:  
; APPLICANT: SHIGENOBU, SHUJI  
; APPLICANT: WATANABE, HIDEKI  
; APPLICANT: HATTORI, MASAHIRA  
; APPLICANT: SAKAKI, YOSHIYUKI  
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
; FILE REFERENCE: 081356/0159  
; CURRENT APPLICATION NUMBER: US/09/790,988  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: JP2000-107160  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 640681  
; TYPE: DNA  
; ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 5.6%; Score 73.4; DB 10; Length 640681;  
Best Local Similarity 55.8%; Pred. No. 4.3e-09;  
Matches 140; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 624 GGCCCGTTTTCTTCAGAGGGGTGGCGACTGTGTGACCAAGTTGTTTAAATATTGTGGAGC 683  
Db 213389 GACCTGGACACTTTAGAGCGGTTACGACAATTATTACTAAGTTATTAACTTATACAAC 213330  
QY 684 CAGATGTGGCTGTGTTCGGTAAAGAGGATTATCAGCAGTGGCGCTTATTCAGAGGATGG 743  
Db 213329 CAGATTTTGCAATTTTGGAGAAAAAGATTATCAACAATTATTAAATATAAAATCCTTG 213270  
QY 744 TTCGAGATCTTGATTTTCCATAAAAGTGATAGGTGCTGAAATAACACGCTGATAATGATG 803  
Db 213269 TAAAGGAGTTGAATTATATATGATAAAATAATTAGCTTACCTACATACGACTAAAAAATG 213210  
QY 804 GCCTGGCAATGAGTTCACGTAATGTGCACCTTTCACCTGAAGAGAGGGGAAAGGCACCTAT 863  
Db 213209 GATTAGCTTTAAGTTCAGAAATAATTATTAAAGTTCCCAAGAAATGAATAGCACCTT 213150  
QY 864 CAATAAATAAA 874  
Db 213149 ATTTATATAA 213139

RESULT 13  
US-09-738-626-132  
; Sequence 132, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162

```

; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 132
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-132

```

```
Query Match      5.4%;      Score 70.8;  DB 10;  Length 837;
Best Local Similarity 55.8%;  Pred. No. 2.5e-10;
Matches 135;  Conservative 0;  Mismatches 107;  Indels 0;  Gaps 0;
```

QY	620	AGCAGGCCCGTTTCTTTCAGAGGGTGGCGACTGTGGTGACGAAGTTGTTTAAATATGTG	6799
Db	358	AGCAGGCCCTGGCCATTTCGATGGTGGCTACCGTGGTGGCGAAGCTGTTCAATTGGTG	417
QY	680	GAGCCAGATGTGGCTGTGTTTCGGTAAGAAGGATTATCAGCACTGGCGGCTTATTCCAGAGG	7399
Db	418	CGCCCTGATCGTGCATATTTTGGACAAAAGATGCTCAGCAGGTTGCGGTGATTCGGCGA	477
QY	740	ATGGTTCGAGATCTTGATTTTCCATAAAAGTGATAGGTGCTGAAATAACACGTGATAAT	7999
Db	478	TTGGTTGCCGATCTAGACATTCCTCCGTGGAGATTCGTCCCGTCCGATTATTTCGTGGCGCC	537
QY	800	GATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTTCACTGAAGAGAGGGAAGGCA	8599
Db	538	GATGGCTTAGCCGAATCCAGCCGCAATCAACGTCTTTCTGCGGATCAGCGAGCGCAAGCT	597
QY	860	CT 861	
Db	598	CT 599	

## RESULT 14

```

US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

Query Match	5.4%;	Score 70.8;	DB 10;	Length 3309400;
Best Local Similarity	55.8%;	Pred. No. 8.4e-08;		
Matches 135: Conservative	0;	Mismatches 107;	Indels 0;	Gaps 0;

QY	620	AGCAGGCCGTTTCTTCTCAGAGGGTGGCGACTGTGGTGACGAAAGTTGTTTAATATTGTG	679
Db	126832	AGCAGGCCCTGGCCATTTCGATGGTGTGGCTACCGTGGTGGCGAAGCTGTTCAATTTCGTG	126773
QY	680	GAGCCAGATGTGGCTGTGTTTCGGTAAGAAAGGATATCAGCAGTGGCGGCTTATTTCAGAGG	739
Db	126772	CGCCCTGATCGTGCAATTTTGGACAAAAGATGCTCAGCAGGTTGCGGTGATTCGGCGA	126713
QY	740	ATGTTTCGAGATCTTGATTTTTCCATAAAAGTCATAGGTGCTGAAATAAACACGTGATAAT	799
Db	126712	TTGGTTGCCGATCTAGACATTCCTCGTGGAGATTCGTCCCGTTCGGATTATTCGTGGCGCC	126653
QY	800	GATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTACCTGAAGAGAGGGAAAAGGCA	859
Db	126652	GATGGCTTAGCCGAATCCAGCCGCAATCAACGTCCTTTCTCGGATCAGCGAGCGCAAGCT	126593
QY	860	CT 861	
Db	126592	CT 126591	

RESULT 15

```

US-09-974-300-5461
; Sequence 5461, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5461
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(501)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-5461

```

Query Match	5.1%;	Score 67.2;	DB 10;	Length 501;
Best Local Similarity	53.9%;	Pred. No. 2.2e-09;		
Matches 138; Conservative	0;	Mismatches 118;	Indels	0;
Gaps	0;			

[illegible]

QY 845 GAGACGGAAAGGCAC 860  
||| | | | |  
Db 467 GAACGCCAAGCCGCAC 482



Search completed: December 16, 2003, 01:41:46  
Job time : 501 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 21:30:25 ; Search time 97 Seconds

(without alignments)  
5988.250 Million cell updates/sec

Title: US-10-033-269-8

Perfect score: 1316

Sequence: 1 tcggcacgaggtttctcag.....tggcattgctatttgtagg 1316

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	118.8	9.0	9381	4	US-09-453-702B-7
2	108	8.2	885	4	US-09-107-532A-1499
3	99.8	7.6	867	4	US-09-134-001C-1484
4	93.8	7.1	852	4	US-09-328-352-3790
C 5	91.8	7.0	579	4	US-09-252-991A-10716
C 6	91.8	7.0	891	4	US-09-252-991A-10624
7	91.8	7.0	909	4	US-09-252-991A-10450
8	91.8	7.0	1422	4	US-09-252-991A-10234
C 9	79	6.0	3717	4	US-09-221-017B-1095
10	70.8	5.4	2164	3	US-09-318-794A-1
11	70.8	5.4	2164	3	US-09-318-793A-3
C 12	69.2	5.3	4403765	3	US-09-103-840A-2
C 13	69.2	5.3	4411529	3	US-09-103-840A-1
14	56.4	4.3	1926	4	US-09-249-585A-4
15	56.4	4.3	1931	2	US-09-130-114-2
C 16	54.4	4.1	7218	1	US-08-232-463-14
C 17	54.2	4.1	619	3	US-08-998-416-842
18	53	4.0	7218	1	US-08-232-463-14
C 19	51	3.9	390	3	US-09-197-649-7
C 20	49.2	3.7	468	4	US-09-252-991A-11520
C 21	49.2	3.7	1419	4	US-09-252-991A-11636
22	49.2	3.7	1587	4	US-09-252-991A-11720
23	49.2	3.7	2304	4	US-09-252-991A-11802
24	49	3.7	390	3	US-09-197-649-7
25	47.8	3.6	465	4	US-09-252-991A-12201
26	47.8	3.6	936	4	US-09-252-991A-12266
27	47.8	3.6	4884	4	US-09-252-991A-12126

C 28	47.8	3.6	4884	4	US-09-252-991A-12292	Sequence 12292, A
C 29	47	3.6	1476	4	US-09-434-288-12	Sequence 12, Appl
C 30	46.8	3.6	1749	4	US-09-450-072-58	Sequence 58, Appl
C 31	46.8	3.6	1749	4	US-09-351-348-58	Sequence 58, Appl
32	46.4	3.5	4674	4	US-09-410-551B-26	Sequence 26, Appl
33	46.4	3.5	4737	4	US-09-410-551B-30	Sequence 30, Appl
C 34	46	3.5	1749	4	US-09-450-072-56	Sequence 56, Appl
C 35	46	3.5	1749	4	US-09-351-348-56	Sequence 56, Appl
36	45.6	3.5	1289	2	US-08-344-833-1	Sequence 1, Appl
37	45.6	3.5	4190	3	US-08-924-345-1	Sequence 1, Appl
38	45.4	3.4	1458	3	US-09-105-537-9	Sequence 9, Appl
39	45.4	3.4	1565	3	US-09-320-878-24	Sequence 24, Appl
40	45.4	3.4	1565	4	US-09-141-908-22	Sequence 22, Appl
41	45.4	3.4	1565	4	US-09-657-440-24	Sequence 24, Appl
C 42	45.4	3.4	13613	3	US-09-105-537-3	Sequence 3, Appl
C 43	45	3.4	933	3	US-09-105-390-43	Sequence 43, Appl
C 44	45	3.4	1008	3	US-09-105-390-59	Sequence 59, Appl
C 45	45	3.4	2810	3	US-09-105-390-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-09-453-702B-7/c

; Sequence 7, Application US/09453702B

; Patent No. 6365723

; GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.

Burland, Valerie

Perna, Nicole T.

Plunkett, Guy

Welch, Rod

TITLE OF INVENTION: No. 6365723e1 Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 9381

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Query Match

Best Local Similarity 9.0%; Score 118.8; DB 4; Length 9381;

48.3%; Pred. No. 1.2e-22;

```
Matches 427; Conservative 0; Mismatches 397; Indels 60; Gaps 1;
QY 239 CGCTCGATCGGGCCCGCAGGCAAGCTCATTTGGGCTGGTCCCCACCATGGGCTTCCCTCCAC 298
Db 1111 CGCCGCTGCTGATGGAAGCAAGCGCGCTGGTACCTACCATGGGCAACCTGCAT 1052
QY 299 GCGGGCCACCTCTCGCTCGTGGCCCGCAGGCGCCCAACTCTCCGACGTCGTCGCCGTCTCC 358
Db 1051 GATGGCCACATGAAGCTGGTCGACGAAGCCAAAGCCCGCGCGATGGTCTGCTCAGT 992
QY 359 ATCTACGTCAACCCGGGCGCAATTCCGCCCCACGGAGGACCTCTCCACCTACCCCTCCGAC 418
Db 991 ATTTTCGTTAAACCCGATGCAGTTCGACCGCCCGGAGGACCTGGCTCGGTACCCACGCACC 932
QY 419 TTCGACGGCGAGCTAAAGAAACTCGCGTCCGTTCCCGGGCGCGTCGACGTCGTTTCCAT 478
Db 931 TTGCAGGAAGACTGCGAAAGCT----- 909
QY 479 CCCCCTAATTGTACGATTACGGGAAGAACGGTGGTGGTGACGTGGCAGAGGCTGGTGA 538
Db 908 -----GAACAAAGCGCAAGTGGAATCTGGTTTTCGCGCCCTCG 872
QY 539 ATGTTGTCGCTGAGAGTGGTCCGGGACGAAAGTTGGTGAGGTTGAGAAAGCTG 598
Db 871 GTAAAGAGATCTACCCGAACGGTACTGAAACCCATACTTACGTGACGTTCCAGGCTT 812
QY 599 GAATTGGGGCTGTGTGGGAAGAGCAGCGCCGCTTTCTTCAGAGGGTGGCGACTGTGGTG 658
Db 811 TCGACCATGCTGGAAGCGCCAGCCGCGGGGCAATTTTCGCGCGTTCGACTATCGTC 752
QY 659 ACGAAGTGTGTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAG 718
Db 751 AGCAAGCTGTTAAACCTGGTCCAGCCGACATGCCTGCTTCGGTGAAGAGGACTTTCAG 692
QY 719 CAGTGGCGGCTTATTCAGAGGATGTTTCGAGATCTTGATTTTCCATAAAAGTATAGGT 778
Db 691 CAACTGGCGTGATCCGCAAAATGTTGCTGATATGGCTTTGATATTGAGATTGTCGGT 632
QY 779 GCTGAAATAACACGTCGATTAATGATGGCTGGCAATGAGTTCAGTAATGTGCACCTTTCA 838
Db 631 GTGCCAATTATCGCGCCCAAGACGGTCTGGCACTGAGTTCGCGTAACGGTTATCTGACG 572
QY 839 CCTGAAGAGAGGAAAGGCACCTATCAATAAATAATCAATTGTTAAGAGCTAAATCAGCA 898
Db 571 GCAGAAACACGCAAAATTGGCGCCCGTCTGTACAAAGTTTAAAGTTCGATTGCCGACAAA 512
QY 899 GCAGGATGGTCAGGTGCATTTGTGAGAAAGTTGACAAATTTGGTTCATCCAAAGTGTACT 958
Db 511 TTGACAGCTGGCGAAACGGGATCTCGATGAAATATTGCTATTGCGGGCAAGAACTGAAT 452
QY 959 GATGCTGGTGAAGGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAGTG 1018
Db 451 GAAAAAGGCTTCGCGCGCGATGATATTTCAGATTTCGCGATGCCGACACATTGCTGGAAGTC 392
QY 1019 GAACAGATCAAGAGTCTCTGCTCTTCTGTTGCTGCTGCTGTTGGCAAAGTCAGGCTT 1078
Db 391 TCTGAAACAGCAACCGGCAGTAATTCTGGTAGCCGCGCTGGCTTGGCGATGCTCGCCTG 332
QY 1079 ATAGACAAATGGAATCAACTTGTCAATGAATGTTTGTATCTAA 1122
Db 331 ATCGAACAAATAATGGTCGAGCTGGCGTAATACTTATCTATA 288
```

```
RESULT 2
US-09-107-532A-1499
; Sequence 1499, Application US/09107532A
; Patent No. 6583275
;
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
```

```
;
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1499:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...885
; SEQUENCE DESCRIPTION: SEQ ID NO: 1499:
US-09-107-532A-1499
```

```
Query Match 8.2%; Score 108; DB 4; Length 885;
Best Local Similarity 52.4%; Pred. No. 3.6e-20;
Matches 262; Conservative 0; Mismatches 235; Indels 3; Gaps 1;
QY 605 GGGCTGTGTGGGAAGAGCAGCGCCCGTTTCTTCAGAGGGGTGGCGACTGTGTTGACGAAG 664
Db 379 GGTATTATGTGGAGCTAGTAGGCCAATACATTTAGAGGGGTTTGCACTGTTGTAAGTAA 438
QY 665 TTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGG 724
Db 439 TTATTTAACATAATACCAGCAGATAGAGCTTATTTTGGAGAAAAGGATGCAACAGTTA 498
QY 725 CGGCTTATTCAGAGGATGGTTCGAGATCTTGATTTTCCATAAAAGTATAGGTGCTGAA 784
Db 499 GCTGTAATTAAAGAAATGGTTAGGACCTTAATATTGATATAGACGTTGTAGGGGTCCA 558
QY 785 ATAACACGTGATATGATGGCTGGCAATGAGTTCACGTAATGTGCACCTTTCACCTGAA 844
Db 559 ATTATAAGGAAGAGGATGGACTAGCTAAAAAGTTCTAGAAAACACATACTTAAGTTAG 618
QY 845 GAGAGGGAAGAGGCACTATCAATAAATAAATCATTTGTTAAGAGCTAAATCAGCAGCAG 904
Db 619 GAAAGATCTTCAGCTACTATTTTAAATAAATCTTTAACTTTAGCTAGGAAGCTTTAAAT 678
QY 905 GATGGTCAGGTGCATTGTGAGAAGTTGACAAATTTGGTCATCCAAAGT---TACTGAT 961
Db 679 AATGGAGAGAGACAGACTTAAATAAATAATTGAGATAATAAGTAAAAATATAAATACTTAT 738
QY 962 GCTGGTGAAGATCGATTATGCTGAGATTGTTGATCAAAATAATTTGGAGAAAAGTGGAA 1021
```







QY 308 CTCTCGCTCGTGGCCAGCCCGCCAACTCTCCGACGTCTGTCGCCGCTCTCCATCTACGTC 367  
| | | | |  
Db 682 GCCGCGCTGGTAAGAACCCGCGAAACGCGCGACTTCGTGGTCGTGACGATCTTCGTC 741  
| | | | |  
QY 368 AACCCGGGCCAATTGCCCCCACCAGGAGGACCTCTCCACCTACCCCTCCGACTTCGACGGC 427  
| | | | |  
Db 742 AACCCGCTGCAATTGCGCCCCCAGCGAAGACCTCGACAAATACCCGCGGACCTTCGCAGCC 801  
| | | | |  
QY 428 GAC 430  
| | | | |  
Db 802 GAC 804

RESULT 9  
US-09-221-017B-1095/c  
; Sequence 1095, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221,017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Monroy, Gladys H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 1095:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3717 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: UNKNOWN  
; ORIGINAL SOURCE:  
; ORGANISM: PORYPHYROMONAS GINGIVALIS  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1...3717  
US-09-221-017B-1095

Query Match 6.0%; Score 79; DB 4; Length 3717;  
Best Local Similarity 56.3%; Pred.No. 7.8e-12;  
Matches 148; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
  
QY 605 GGGCTGTGTGGGAAGAGCAGCGCCCGTTTTCTTCAGAGGGGTGGGACTGTGGTGACGAAG 664  
| | | | |  
Db 2265 GTGATGGAAGGCAAGCATCGCCCCCGACACTTCAATGGCGTGGCGCAGGTGGTAAGCAA 2206  
| | | | |  
QY 665 TTGTTTAATATTGTGGAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGG 724  
| | | | |  
Db 2205 CTCTTCATGATGGTGGAGCCCGGACAAGGCATATTTCCGAGAAAAGGATTTCAGCAATC 2146  
| | | | |  
QY 725 CGGCTTAATTCAGAGGATGGTTCGAGATCTTGATTTTTCCATAAAAGTAGGTGCTGAA 784  
| | | | |  
Db 2145 GCCGTTATCCGACGATGGTGAACCTGTTGGCTTACCCGTTACGATCGTGGCCTGTCT 2086  
| | | | |  
QY 785 ATAACACGTGATAATGATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTCACCTGAA 844  
| | | | |  
Db 2085 ATCATCCGTGAAGAGGACGGGCTTGCACTCAGTAGCCGCAACGTCCGCTCGGCACCGAA 2026  
| | | | |  
QY 845 GAGAGGGGAAAAGGCACATATCAAT 867  
| | | | |  
Db 2025 GAGCGTGCCATAGCCCCCATCCAT 2003  
| | | | |

RESULT 10  
US-09-318-794A-1  
; Sequence 1, Application US/09318794A  
; Patent No. 6177264  
; GENERAL INFORMATION:  
; APPLICANT: DEGUSSA AKTIENGESSELLSCHAFT  
; TITLE OF INVENTION: METHOD FOR THE FERMENTATIVE PRODUCTION OF D-PANTOTHENIC  
; TITLE OF INVENTION: ACID USING CORYNEFORM BACTERIA  
; FILE REFERENCE: eggeling  
; CURRENT APPLICATION NUMBER: US/09/318,794A  
; PRIOR FILING DATE: 1999-05-26  
; PRIOR APPLICATION NUMBER: DE 198 55 312.9  
; PRIOR FILING DATE: 1998-12-01  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2164  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-318-794A-1

Query Match 5.4%; Score 70.8; DB 3; Length 2164;  
Best Local Similarity 55.8%; Pred.No. 1.1e-09;  
Matches 135; Conservative 0; Mismatches 107; Indels 0; Gaps 0;  
  
QY 620 AGCAGGCCCGTTTTCTTCAGAGGGGTGGCGACTGTGGTGACGAAGTGTGTTAATATTGTG 679  
| | | | |  
Db 1523 AGCAGGCCCTGGCCCATTTTCGATGGTGTGGTACCGTGGTGGCGAAGCTGTTCAATTGGTG 1582  
| | | | |  
QY 680 GAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGGCGGCTTATTTCAGAGG 739  
| | | | |  
Db 1583 CGCCCTGATCGTGCATATTTTGGACAAAAGATGCTCAGCAGGTTGCGGTGATTTCGGCGA 1642  
| | | | |  
QY 740 ATGGTTCGAGATCTTGATTTTTCCATAAAAGTAGGTGCTGAATAACACGTGATAAT 799  
| | | | |  
Db 1643 TTGGTTGCCGATCTAGACATTCCTCGTGAGATTCTGTCCTCCGTTCCGATTATTCTGGCGCC 1702  
| | | | |  
QY 800 GATGGCCTGGCAATGAGTTACGTAATGTGCACCTTTTCACCTGAAGAGAGGAAAGGCA 859  
| | | | |  
Db 1703 GATGGCTTAGCCGAATCCAGCCGCAATCAACGTCCTTTCTCGCGATCAGCGAGCGCAAGCT 1762  
| | | | |  
QY 860 CT 861  
| |  
Db 1763 CT 1764  
| |

RESULT 11  
US-09-318-793A-3



```
; Sequence 3, Application US/09318793A
; Patent No. 6184007
; GENERAL INFORMATION:
; APPLICANT: Dusch, Nicole
; APPLICANT: Kalinowski, Jorn
; APPLICANT: Puhler, Alfred
; TITLE OF INVENTION: METHOD FOR THE FERMENTATIVE PRODUCTION OF D-PANTOTHENIC
; FILE REFERENCE: 21123/260204
; CURRENT APPLICATION NUMBER: US/09/318,793A
; CURRENT FILING DATE: 1999-05-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2164
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (351)..(1163)
; NAME/KEY: CDS
; LOCATION: (1166)..(2002)
US-09-318-793A-3
```

```
Query Match 5.4%; Score 70.8; DB 3; Length 2164;
Best Local Similarity 55.8%; Pred. No. 1.1e-09;
Matches 135; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 620 AGCAGGCCCGTTTCTTCAGAGGGGTGGCGACTGTGGTGACGAAGTTGTTAATATTGIG 679
Db 1523 AGCAGGCCCTGGCCATTTCGATGGTGCTACCGTGGTGGCGAAGCTGTTCAATTGGTG 1582

QY 680 GAGCCAGATGTGGCTGTGTTCCGTAAGAAGGATTATCAGCAGTGGCGGCTTATTCAGAGG 739
Db 1583 CGCCCTGATCGTGCAATATTTTGGACAAAAAGATGCTCAGCAGGTTCGGTGATTCGGCGA 1642

QY 740 ATGGTTCGAGATCTTGATTTTCCATAAAAAGTGATAGGTGCTGAATAACACGTGATAAT 799
Db 1643 TTGGTTGCCGATCTAGACATTCCCGTGGAGATTCTGCCGTTCCGATTATTCGTGGCGCC 1702

QY 800 GATGGCCTGGCAATGAGTTCACGTAATGTGCACCTTTTCACCTGAAGAGAGGAAAGGCA 859
Db 1703 GATGGCTTAGCGGAATCCAGCCGCAATCAACGTCTTTCTGCGGATCAGCGAGCGCAAGCT 1762

QY 860 CT 861
Db 1763 CT 1764
```

```
RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
```

```
US-09-103-840A-2

Query Match 5.3%; Score 69.2; DB 3; Length 4403765;
Best Local Similarity 59.0%; Pred. No. 1.4e-07;
Matches 138; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

QY 239 CGCTCGATCGGGCCAGGGCAAGCTCATTGGGTGGTCCCCACCACATGGGCTTCCTCCAC 298
Db 4037206 CGCGCACTGCGACTACCGGCCGCGAGTGATGTTGGTGCCTACTATGGTGCGTGCAC 4037147

QY 299 GCGGGCCACCTCTCGCTCGTGGCCAGGCCCGCCAACTCTCCG---ACGTCGTGCGCGTC 355
Db 4037146 GAAGGCCACCTCGCGTTGGTGGTGGCGCAAGCGGTGCCCGGATCGGTCGTGCTGTG 4037087

QY 356 TCCATCTACGTCAACCCGGGCCAAATTGCGCCCGCCACGAGGACCTCTCCACCTACCCCTCC 415
Db 4037086 TCGATCTTCGTCAACCCCGATGCAATTGCGTGCCGGGGAAGATCTCGACGCCCTATCCCGC 4037027

QY 416 GACTTCGACGGCGACGTAAAGAAAACCTCGCGTCCGTTCCCGGGCGGCGTCGACGTC 469
Db 4037026 ACCCGGACGACGACCTTGGCGCAACTGCGGGCCGAAGCGGTGGAATCGCTTTC 4036973

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 5.3%; Score 69.2; DB 3; Length 4411529;
Best Local Similarity 59.0%; Pred. No. 1.4e-07;
Matches 138; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

QY 239 CGCTCGATCGGGCCAGGGCAAGCTCATTGGGTGGTCCCCACCACATGGGCTTCCTCCAC 298
Db 4045135 CGCGCACTGCGACTACCGGCCGCGAGTGATGTTGGTGCCTACTATGGGTGCGTGCAC 4045076

QY 299 GCGGGCCACCTCTCGCTCGTGGCCCGCCCGCCAACTCTCCG---ACGTCGTGCGCGTC 355
Db 4045075 GAAGGCCACCTCGCGTTGGTGGTGGCGCAAGCGGTGCCCGGATCGGTCGTGCTGTG 4045016

QY 356 TCCATCTACGTCAACCCGGGCCAAATTGCGCCCGCCACGAGGACCTCTCCACCTACCCCTCC 415
Db 4045015 TCGATCTTCGTCAACCCCGATGCAATTGCGTGCCGGGGAAGATCTCGACGCCCTATCCCGC 4044956

QY 416 GACTTCGACGGCGACGTAAAGAAAACCTCGCGTCCGTTCCCGGGCGGCGTCGACGTC 469
Db 4044955 ACCCGGACGACGACCTTGGCGCAACTGCGGGCCGAAGCGGTGGAATCGCTTTC 4044902

RESULT 14
US-09-249-585A-4
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 13:28:44 ; Search time 20 Seconds  
(without alignments)  
1490.615 Million cell updates/sec

Title: US-10-033-269-9  
Perfect score: 1588  
Sequence: 1 MAPAPRVISDKASMRWSRS.....AWFGKVRLIDNMEINLSMNV 310

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	927.5	58.4 313 2 T03924	probable pantoate-
2	639.5	40.3 280 2 E72296	pantoate-beta-alan
3	537	33.8 281 2 F97258	pantoate-beta-alan
4	536	33.8 282 2 E81448	pantoate-beta-alan
5	518.5	32.7 282 2 G70482	pantothenate synth
6	516	32.5 283 2 E64736	pantoate-beta-alan
7	514.5	32.4 285 2 E87517	pantoate-beta-alan
8	514	32.4 283 2 A99646	pantothenate synth
9	514	32.4 283 2 A85497	pantothenate synth
10	513	32.3 283 2 H83860	pantothenate synth
11	504.5	31.8 283 2 F90066	pantoate beta-alan
12	500	31.5 284 2 AG0524	pantoate,beta-alan
13	499	31.4 293 2 H82303	pantoate-beta-alan
14	498	31.4 284 2 AD0413	pantoate-beta-alan
15	497	31.3 283 2 S55486	probable pantoate-
16	497	31.3 283 2 S75604	pantothenate synth
17	494	31.1 286 2 H69671	pantothenate synth
18	489.5	30.8 278 2 G81148	pantoate-beta-alan
19	488	30.7 281 2 F82832	pantoate-beta-alan
20	487	30.7 283 2 G83055	pantoate-beta-alan
21	480.5	30.3 278 2 F81874	probable pantoate-
22	467	29.4 534 2 AI2172	pantothenate synth
23	465	29.3 292 2 AI2984	pantoate-beta-alan
24	465	29.3 345 2 F98298	pantoate-beta-alan
25	457.5	28.8 285 2 A84953	pantoate-beta-alan
26	456	28.7 281 2 G75430	pantoate-beta-alan
27	455.5	28.7 279 2 T47120	pantoate-beta-alan
28	451	28.4 345 2 S48389	probable membrane
29	447	28.1 293 2 AC3451	pantoate-beta-alan

30	423.5	26.7	313	2	F86937	probable pantoate-
31	418	26.3	285	2	AE1312	pantothenate synth
32	415.5	26.2	309	2	C70955	probable panC prot
33	407	25.6	276	2	H71985	pantoate-beta-alan
34	405	25.5	276	2	F64520	pantoate-beta-alan
35	401	25.3	285	2	AE1684	pantothenate synth
36	330	20.8	197	2	T36071	probable pantoate-
37	168.5	10.6	176	2	T36394	probable pantoate-
38	110	6.9	459	2	AC2116	hypothetical prote
39	100	6.3	529	2	E87259	hypothetical prote
40	98.5	6.2	4307	2	T20721	hypothetical prote
41	96.5	6.1	280	2	T36066	hypothetical prote
42	95.5	6.0	550	2	F90407	molybdopterin bios
43	95.5	6.0	2535	2	AC0304	probable hemolysin
44	94.5	6.0	573	2	B70942	hypothetical prote
45	93.5	5.9	423	2	G83573	conserved hypothet

ALIGNMENTS

RESULT 1

T03924

probable pantoate-beta-alanine ligase (EC 6.3.2.1) - rice

C;Species: Oryza sativa (rice)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-Jun-2002

C;Accession: T03924

R;Genschel, U.

submitted to the EMBL Data Library, December 1996

A;Reference number: Z15136

A;Accession: T03924

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-313 <GEN>

A;Cross-references: EMBL:Y10253; PIDN:CAA71303.1

A;Experimental source: cv. Nipponbare

C;Genetics:

A;Gene: panC

C;Superfamily: pantoate-beta-alanine ligase

C;Keywords: coenzyme A biosynthesis; ligase

Query Match 58.4%; Score 927.5; DB 2; Length 313;  
Best local Similarity 59.9%; Pred. No. 2.6e-65;  
Matches 185; Conservative 51; Mismatches 60; Indels 13; Gaps 6;

Qy	5	PRVISDKASMRWSRSMRAQGLIGLVPTMGFLHAGHLSLV--AQARQLSDVVA--VSIY	60
		:     :     :     :     :     :     :     :     :	
Db	7	PEVIRDKAAMRAWSRRRRRAEGKTAVVPTMGYLHQHLSLISAAAAAASADPVAIVVTIY	66
		:     :     :     :     :     :     :     :	
Qy	61	VNPGQFAPTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLY--DYGNKNGGDDVAEAGG	118
		:     :     :     :     :     :     :     :	
Db	67	VNPSQFAPSEDLATYPSDFAGDLRKLAST-GVVDVAVFNPPDLVVRGAGRRGAG---SGG	121
		:     :     :     :     :     :     :     :	
Qy	119	MVSCVE--SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDAVFGKKD	176
		:   : :     :     :     :     :     :     :     :	
Db	122	AISCLEEAAGDGHETWVRVERLEKGLCGASRPVFFRGVATIVSKLFNIEPDVPVFGKKD	181
		:     :     :     :     :     :     :     :	
Qy	177	YQQRLLIQRMVRDLDFSIIKVI GAETRDNDGLAMSSRNVHLSPEEREKALSINKSLRAK	236
		:     :     :     :     :     :     :     :	
Db	182	YQQRVILPYWSGLDFGIEIMGSRNCARTDGLAMNSRNVHLSREECKKALSISRSLVDAR	241
		:     :     :     :     :     :     :     :	
Qy	237	SAAGDGVHCEKLTNLVLIQSVTDAGGRIDYAEIVDONNLEKVEIKSPVFCVAAWFGKV	296
		:   : :     :     :     :     :     :     :     :	
Db	242	TGALKGNTDSKQIKNKIVQTLTETGGQVDYVEIVEQESLVPVEIQIDGPPVICVAAWFGKV	301
		:     :     :     :     :     :     :     :	
Qy	297	RLIDNMEIN	305
		:	
Db	302	RLIDNIEID	310

RESULT 2

E72296

pantoate-beta-alanine ligase - Thermotoga maritima (strain MSB8)





Db 275 N 275

RESULT 5  
G70482  
pantothenate synthetase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 29-Sep-1999  
C:Accession: G70482  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: G70482  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-282 <AQF>  
A:Cross-references: GB:AE000774; NID:g2984324; PIDN:AAC07847.1; PID:g2984326; GB:AE00069  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: panC  
C:Superfamily: pantoate-beta-alanine ligase

Query Match 32.7%; Score 518.5; DB 2; Length 282;  
Best Local Similarity 40.6%; Pred. No. 2.9e-33;  
Matches 123; Conservative 47; Mismatches 106; Indels 27; Gaps 5;

QY 5 PRVSDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP 64  
Db 2 PLLFKKIKDLRNFKNKRCGKBIQVPTMGYLHEGHRQLIKLARMQNDIVVVSIFVNPT 61

QY 65 QFAPTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLVDYGNKGGDVAEAGGMVSCVE 124  
Db 62 QFGEDEDYDRPRDLERDLE--ICEEAGVDVVFAPVEVDRIYPK----- 102

QY 125 SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLNFIVEPDVAVFGKKDYQWRLLIQ 184  
Db 103 ---GYRTKVCVGELGKVLGEFRPGHFDGVATIVVKLFNIVQPNRAYFGEKDYQQLKIE 159

QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRKSAAGDGQV 244  
Db 160 QVVEDLNIPVEIVPPIVREEDGLAYSSRNVLSPERESALSIIKSFLLAEKMIKAGER 219

QY 245 HCEKLTNLV---IQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVVFCAAWFGKVRLLDN 301  
Db 220 DAKRIKEAIRAFIERHPHVKG-VDYVEITDQ-NLNPKETVEKGRILVAVRVGNARLIDN 277

QY 302 MEI 304  
Db 278 WKV 280

RESULT 6  
E64736  
pantoate-beta-alanine ligase (EC 6.3.2.1) - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 03-Jun-2002  
C:Accession: E64736  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: E64736  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-283 <BLAT>  
A:Cross-references: GB:AE000122; GB:U00096; NID:g1786315; PIDN:AAC73244.1; PID:g1786325;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: panC  
C:Superfamily: pantoate-beta-alanine ligase

C;Keywords: coenzyme A biosynthesis; ligase

Query Match 32.5%; Score 516; DB 2; Length 283;  
Best Local Similarity 38.9%; Pred. No. 4.6e-33;  
Matches 118; Conservative 52; Mismatches 109; Indels 24; Gaps 3;

QY 7 VISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNPQGF 66  
Db 3 ILETPLLRQIIRLRMEGRKRVLVPTMGNLHGHMKLVDEAKARADVIVVSIFVNPMPQF 62

QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHP--RNLYDYGNKGGDVAEAGGMVSCVE 124  
Db 63 DRPEDLARYPRTLQEDCEKLNK--RKVDLVFAPSVKIY-----P 100

QY 125 SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLNFIVEPDVAVFGKKDYQWRLLIQ 184  
Db 101 NGTETHYVDVPGGLSTMLEGASRPGHFRGVSTIVSKLFLNLVQPDIIACFGEKDFQQLALIR 160

QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRKSAAGDGQV 244  
Db 161 KMVADMGFDIEIVGVPIIMRAKDGALSSRNGYLTAEQRKIAPGLYKVLSSIIADKIQAGER 220

QY 245 HCEKLTNLVIQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVVFCAAWFGKVRLLDNMEI 304  
Db 221 DLDEIITIAGQELNEKGFRADDIQIRDADTLLVEVSETSKRAVILVAWLGDARLIDNKV 280

QY 305 NLS 307  
Db 281 ELA 283

RESULT 7  
E87517

pantoate-beta-alanine ligase [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: E87517  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87517  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <STO>  
A:Cross-references: GB:AE005673; NID:g13423661; PIDN:AAK24137.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC2166  
C:Superfamily: pantoate-beta-alanine ligase

Query Match 32.4%; Score 514.5; DB 2; Length 285;  
Best Local Similarity 39.7%; Pred. No. 6.1e-33;  
Matches 122; Conservative 49; Mismatches 105; Indels 31; Gaps 6;

QY 4 APRVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP 63  
Db 3 SPIIVRTVAEMREHVRANKAAGQRVAVVPTMGALHEGHLNLRVLAQQHAERVIATVFN 62

QY 64 QGFAPTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLVDYGNKGGDVAEAGGMVSCV 123  
Db 63 KQFAPHEDFDAYPRGEADAADAEKALV--GCDLLFAPNATEMY----- 102

QY 124 ESGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLNFIVEPDVAVFGKKDYQWRLLI 183  
Db 103 --APGFSTLVSVSGVSEPLEGARPPQFFGGVATVVAKLFIQSADVAVFGEKDYQQLQV 160

QY 184 QRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRKSAAGDGQ 243  
Db 161 RRMARDLIPVEIIGAPTARAEDGLALSSRNAYLSAEERAAVALPTAMKAAAQAQGG 220

QY 244 VHCEKLTNLVIQSVTDAG-GRIDYAEIVDQNNLEKVEIKSPV-----VFCVAWFGKVR 297

Db 221 -PIEDAERSAVALQAAGFGQVDYVEIREASDLRLG--PGPIGEASGRILVAAWLGKTR 277  
QY 298 LIDNMEI 304  
Db 278 LIDNMAV 284

RESULT 8  
A99646

panthothenate synthetase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C;Accession: A99646  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: A99646  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-283 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BAB33560.1; PID:g13359593; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: ECs0137  
C;Superfamily: pantoate-beta-alanine ligase

Query Match 32.4%; Score 514; DB 2; Length 283;  
Best Local Similarity 38.9%; Pred. No. 6.7e-33;  
Matches 118; Conservative 52; Mismatches 109; Indels 24; Gaps 3;

QY 7 VISDKASMRWSRSMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGF 66  
Db 3 IIETLPLLRQIIRRLRMEGKRVALVPTMGNLHDGKMLVDEAKARADVVSIFVNPMPQF 62  
QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVDDVVFHP--RNLYDYGNKGGDVAEAGGMVSCVE 124  
Db 63 DRPEDLARYPRTLQEDCEKLNK--RKVDLVFAPSVKEIY-----P 100  
QY 125 SGSGHESWVRVEKLELGLCGKSRPVFRGVATVTKLFNIVEPDVAVFGKKDYQQWRLIQ 184  
Db 101 NGTETHYVDVPGSLTMLEGASRPGHFRGVSTIVSKFLNVLQPDIAACFGEKDFQQLALIR 160  
QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLLRAKSAAGDQV 244  
Db 161 KMVADMGFDEIVGVPIMRAKDGLALSSRNGYLTAEQRKIAPGLYKVLSSIAADKLQAGER 220  
QY 245 HCEKLTNLVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVFCVAAWFGKVRLLDNMEI 304  
Db 221 DLDEITAIAGQELNEKGFRADDIQIRDADTLLEVSETSKRAVILVAWLGARLLIDNKMV 280  
QY 305 NLS 307  
Db 281 ELA 283

RESULT 9  
A85497

panthothenate synthetase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: A85497  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: A85497  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-283 <STO>

A;Cross-references: GB:AE005174; NID:g12512848; PIDN:AAG54437.1; GSPDB:GN00145; UWGP:2  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: panC  
C;Superfamily: pantoate-beta-alanine ligase

Query Match 32.4%; Score 514; DB 2; Length 283;  
Best Local Similarity 38.9%; Pred. No. 6.7e-33;  
Matches 118; Conservative 52; Mismatches 109; Indels 24; Gaps 3;

QY 7 VISDKASMRWSRSMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGF 66  
Db 3 IIETLPLLRQIIRRLRMEGKRVALVPTMGNLHDGKMLVDEAKARADVVSIFVNPMPQF 62  
QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVDDVVFHP--RNLYDYGNKGGDVAEAGGMVSCVE 124  
Db 63 DRPEDLARYPRTLQEDCEKLNK--RKVDLVFAPSVKEIY-----P 100  
QY 125 SGSGHESWVRVEKLELGLCGKSRPVFRGVATVTKLFNIVEPDVAVFGKKDYQQWRLIQ 184  
Db 101 NGTETHYVDVPGSLTMLEGASRPGHFRGVSTIVSKFLNVLQPDIAACFGEKDFQQLALIR 160  
QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLLRAKSAAGDQV 244  
Db 161 KMVADMGFDEIVGVPIMRAKDGLALSSRNGYLTAEQRKIAPGLYKVLSSIAADKLQAGER 220  
QY 245 HCEKLTNLVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVFCVAAWFGKVRLLDNMEI 304  
Db 221 DLDEITAIAGQELNEKGFRADDIQIRDADTLLEVSETSKRAVILVAWLGARLLIDNKMV 280  
QY 305 NLS 307  
Db 281 ELA 283

RESULT 10  
H83860

panthothenate synthetase panC [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: H83860  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hi  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans ar  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: H83860  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-283 <STO>  
A;Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05407.1; GSPDB:GN  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: panC  
C;Superfamily: pantoate-beta-alanine ligase

Query Match 32.3%; Score 513; DB 2; Length 283;  
Best Local Similarity 40.3%; Pred. No. 8e-33;  
Matches 122; Conservative 46; Mismatches 95; Indels 40; Gaps 6;

QY 8 ISDKASMRWSRSMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQQA 67  
Db 7 ISD---LQAQLRIEREQKRSVGFVPTMGYLHEGHLSLVKRAKEEHTVVSIFVNPLOFG 63  
QY 68 PTEDLSTYPSDFDGDVKKLASVPGGVDDVVFHPNRNLYDYGNKGGDVAEAGGMVSCVESGS 127  
Db 64 AGEDLDTPRDFARD-EQLAAE-GVDILFYP-----S 94  
QY 128 GHESWVRVEKLELG-----LCGKSRPVFRGVATVTKLFNIVEPDVAVFGKKDYQQW 180  
Db 95 TDEMYPRPASVRLKVTQGVDDVLCASRPGHFDGVVTVLKFLHVEPDAAAYFGLKDAQV 154  
QY 181 RLQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLLRAKSAAG 240



Db 155 AVITNMVEDLNVGVQIVPCATVREVDGLAKSSRNVRLSEKERKEAPGLYQSLLAGREALD 214

QY 241 DQGVHCEKLTNLVIOQVTD-AGGRIDYAEIVDQNNLEKVBQIKSPVWFCVAAWFGKVRLLI 299

Db 215 AGEKDAAVIRERIRQSLERLTGRIDYVEVLSYPRLOKIERIEETVILAVAYQFENARLI 274

QY 300 DNM 302

Db 275 DNL 277

RESULT 11

F90066

pantoate beta-alanine ligase [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C;Accession: F90066

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.

A;Reference number: A89758; PMID:21311952; PMID:11418146

A;Accession: F90066

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-283 <KUR>

A;Cross-references: GB:BA000018; PID:g13702555; PIDN:BAB43696.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: panC

C;Superfamily: pantoate-beta-alanine ligase

Query Match 31.8%; Score 504.5; DB 2; Length 283;

Best Local Similarity 37.2%; Pred. No. 3.7e-32;

Matches 113; Conservative 58; Mismatches 102; Indels 31; Gaps 5;

QY 6 RVISDKASMRSWSRMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQG 65

Db 3 KLITTVKEMQHIVKAAKRS GTTIGFIPTMGALHDGHLTMVRESVSTNDITVVSFVNPLQ 62

QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVDVVFHPNLYDYGKNGGGDVAEAGGMVSCVES 125

Db 63 FGPNEFDAYPRQIDKDELVSEV--GADIVFHPAVEDIYPGELGIDV-KVGPLADVLE- 118

QY 126 GSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKDYQQWRLLIQR 185

Db 119 -----GAKRPGHFDGVVTVYVVKLFNIVMPDYAYFGKDAQQLAIVEQ 160

QY 186 MVRDLDFSIVKIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLPRAKSAAGDGQVH 245

Db 161 MVKDFNHAVEIIGDIVREADGLAKSSRNVLTEQERQEAHVLSKSLLAQALYQDG--- 217

QY 246 CEKLTNLVIOQVTD-----AGGRIDYAEIVDQNNLEKVEQIKSPVWFCVAAWFGKVRLLID 300

Db 218 -ERQSKVIIDRVTEYLESHISGRIEEVAVSYVQLVEQHEITGRIFISLAVKFSKARLLID 276

QY 301 NMEI 304

Db 277 NIII 280

RESULT 12

AG0524

pantoate,beta-alanine ligase [imported] - Salmonella enterica subsp. enterica serovar Ty

C;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AG0524

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, , S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AG0524

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-284 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD01335.1; PID:g16501463; GSPDB:GN00176

C;Genetics:

A;Gene: STY0199

C;Superfamily: pantoate-beta-alanine ligase

Query Match 31.5%; Score 500; DB 2; Length 284;

Best Local Similarity 37.4%; Pred. No. 8.4e-32;

Matches 114; Conservative 57; Mismatches 106; Indels 28; Gaps 5;

QY 7 VISDKASMRSWSRMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGQF 66

Db 3 I IETLPLLRQHTRRLRQEGKRVALVPTMGNLHDGHMKLVDEAKARADVIVSIFVNPMPQF 62

QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVDVVFHP--RNLYDYGKNGGGDVAEAGGMVSCVE 124

Db 63 DRPDDLVRYPRTLQEDCEKLNK--RKVDYVFAPAVEEIIYPQGLEG----- 105

QY 125 GSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKDYQQWRLLIQ 184

Db 106 -----QTYVDVPGLSTMLEGTSRPGHFRGVSTIVSKLENLIQPDIACFGEKDFQQLALIR 160

QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSL--LRAKSAAGDG 242

Db 161 KMVADMGYDIEIVGVPIIRAKOGLALSSRNAYLTAEQRKIAPGLYNVWNSIAEKLIAGNR 220

QY 243 QVHCEKLTNLVIOQVTDAGGRIDYAEIVDQNNLEKVEQIKSPVWFCVAAWFGKVRLLIDNM 302

Db 221 EL--QEIIIAEQLNEKGFRADDIQIRDADTLLLETTETSKRAVILAAAWLGQARLLDNQ 278

QY 303 EINLS 307

Db 279 SVTLA 283

RESULT 13

H82303

pantoate-beta-alanine ligase VC0591 [imported] - Vibrio cholerae (strain N16961 serogr

C;Species: Vibrio cholerae

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: H82303

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: H82303

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-293 <HEI>

A;Cross-references: GB:AE004144; GB:AE003852; NID:g9655017; PIDN:AAF93758.1; GSPDB:GN0

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC0591

A;Map position: 1

C;Superfamily: pantoate-beta-alanine ligase

Query Match 31.4%; Score 499; DB 2; Length 293;

Best Local Similarity 37.0%; Pred. No. 1e-31;

Matches 121; Conservative 54; Mismatches 80; Indels 72; Gaps 7;

QY 6 RVISDKASMRSWSRMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQG 65

Db 2 QVFADIAVLRQIKQIKREGRRVAFVPTMGNLHEGHLTLVRKARELADVIVVVSIFVNPQM 61

QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVDVVF--HPRNLYDYGKNGGGDVAEAGGMVSCV 123

Db 62 FDRAEDLNKYPRTLEEDLSKNG--EGVDLVLTPTPETMYPQGLD----- 104  
QY 124 ESGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLI 183  
Db 105 -----KQTFVEVPGLSYMLEGASRPGHFRGVATIVTKLFNIVDPVACFGEKDFQQLAVI 159  
QY 184 QRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLRAKSAAGDQ 243  
Db 160 QMWVEDLCMDIEIVGVATVRELDGLAMSSRNLLTLDERQAPVPLARTMRWISSAI---- 215  
QY 244 VHCEKLTNLVVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKS-----PV- 285  
Db 216 -----RGRDDYPSIIE----DAVDQLRAADLEPDEFIRDARTLLPIS 255  
QY 286 -----VFCVAAWFGKVRLLDNMEINL 306  
Db 256 SESKQAVILMSAFLGKVRLLDNQVLDL 282

## RESULT 14

AD0413  
pantoate-beta-alanine ligase (EC 6.3.2.1) [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 03-Jun-2002  
C:Accession: AD0413  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AD0413  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-284 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92632.1; PID:G15981329; GSPDB:GN00175  
C:Genetics:  
A:Gene: panC  
C:Superfamily: pantoate-beta-alanine ligase  
C:Keywords: coenzyme A biosynthesis; ligase

Query Match 31.4%; Score 498; DB 2; Length 284;  
Best Local Similarity 39.7%; Pred. No. 1.2e-31;  
Matches 120; Conservative 48; Mismatches 110; Indels 24; Gaps 6;

QY 7 VTSKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQF 66  
Db 3 ILETPLLRQQIRRRWRQEGKRIALVPTMGNLHEGHMTLVDEAKTRADVVTIVFNPLQF 62  
QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVVDFHPRNLYDGKNGGGDVAEAGGMVSCVESG 126  
Db 63 ERPDLAHYPTLQEDCEKLTUR--HGADLVFAP-----AAADIYPAG-----LEK- 105  
QY 127 SGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLI 186  
Db 106 ---QTYVDVPALSTILEGASRPGHFRGVSTIVSKLFNLIQPDVACFGEKDYQQLALIRKM 162  
QY 187 VRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSL--LEAKSAAGDQV 244  
Db 163 VADMGYDINIVGVPTVRAKDLGLALSSRNGYLTEEEERQIAPQLSKIMWALAEKMGALGERQI 222  
QY 245 HCEKLTNLVVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVVFCAAWFGKVRLLDNMEI 304  
Db 223 --DALLEEAAQALLRVGFTPELFIKDAETLQPLTVDSQQAVILMAAWLGAARLIDNQLV 280  
QY 305 NL 306  
Db 281 DL 282

RESULT 15  
S55486

probable pantoate-beta-alanine ligase (EC 6.3.2.1) - fission yeast (Schizosaccharomyce  
N;Alternate names: protein AcSH10.08c  
C:Species: Schizosaccharomyces pombe  
C:Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 03-Jun-2002  
C:Accession: S55486; T38972  
R:Connor, R.; Churcher, C.M.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S55479  
A:Accession: T38972  
A:Molecule type: DNA  
A:Residues: 1-283 <CON>  
A:Cross-references: EMBL:Z49811; NID:g854599; PIDN:CAA89958.1; PID:g854607  
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: Z21821  
A:Accession: T38972  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-283 <CO2>  
A:Cross-references: EMBL:Z49811; PIDN:CAA89958.1; GSPDB:GN00066; SPDB:SPACSH10.08c  
A:Experimental source: strain 972h-; cosmid c5H10  
C:Genetics:  
A:Gene: SPDB:SPACSH10.08c  
A:Map position: 1  
C:Superfamily: pantoate-beta-alanine ligase  
C:Keywords: coenzyme A biosynthesis; ligase

Query Match 31.3%; Score 497; DB 2; Length 283;  
Best Local Similarity 40.1%; Pred. No. 1.4e-31;  
Matches 122; Conservative 44; Mismatches 114; Indels 24; Gaps 3;

QY 6 RVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQ 65  
Db 2 QVLKEKLLIHQQVDNWRKDGNRIFAFTMGNLHEGHFSLVREAKRHAEKVVVSIYVNPQ 61  
QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVVDFHPRNLYDGKNGGGDVAEAGGMVSCV 123  
Db 62 FNNPQDLLLYPRTMDQCSQLQNL--GVDLVYAPTVEELY----- 99  
QY 124 ESGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLI 183  
Db 100 PEGSQDITFVDVPKLSLTMLEGASRPGHFRGVTVVSKLFHIVNPDVACFGEKDFQQAII 159  
QY 184 QRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLRAKSAAGDQ 243  
Db 160 KKMVRDLNFFIEIIQVPIVRADDGLALSSRNGYLTSEERKIAPNLYKILKLAQELSGN 219  
QY 244 VHCEKLTNLVVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVVFCAAWFGKVRLLDNME 303  
Db 220 GDLKLIATNTLSRCRFIPDQLEICDSTTLEPFTAGTKNVVILAAAWLGAARLIDNIO 279  
QY 304 INLS 307  
Db 280 TTIN 283

Search completed: December 15, 2003, 13:34:17  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 13:29:19 ; Search time 17 Seconds  
(without alignments)  
857.546 Million cell updates/sec

Title: US-10-033-269-9

Perfect score: 1588

Sequence: 1 MAPAPRVISDKASMRWSRS.....AWFGKVRLLDNMEINLSMNV 310

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1286	81.0	307	1 PANC_LOTJA	O24035 lotus japon
2	1101	69.3	310	1 PANC_ARATH	Q9fkb3 arabidopsis
3	927.5	58.4	313	1 PANC_ORYSA	O24210 oryza sativ
4	639.5	40.3	280	1 PANC_THEMA	Q9x0g6 thermotoga
5	631.5	39.8	280	1 PANC_THENE	O86953 thermotoga
6	545	34.3	283	1 PANC_BRAJA	Q9amr9 bradyrhizob
7	537	33.8	281	1 PANC_CLOAB	Q97f38 clostridium
8	536	33.8	282	1 PANC_CAMJE	Q9pik2 campylobact
9	518.5	32.7	282	1 PANC_AQAJE	O67891 aquifex aeo
10	516	32.5	283	1 PANC_ECOLI	P31663 escherichia
11	514.5	32.4	285	1 PANC_CAUCR	Q9a6c8 caulobacter
12	514	32.4	283	1 PANC_ECO57	Q9x930 escherichia
13	513	32.3	283	1 PANC_BACHD	Q9kc86 bacillus ha
14	506.5	31.9	281	1 PANC_SHEON	Q8eih0 shewanella
15	504.5	31.8	283	1 PANC_STAAM	Q99r39 staphylococ
16	504.5	31.8	283	1 PANC_STAAM	Q8nnu2 staphylococ
17	504	31.7	284	1 PANC_SALTY	Q8zrr1 salmonella
18	500	31.5	284	1 PANC_SALTI	Q8z9d3 salmonella
19	499	31.4	293	1 PANC_VIBCH	Q9kud1 vibrio chol
20	498	31.4	284	1 PANC_YERPE	Q8zbb7 yersinia pe
21	497	31.3	283	1 PANC_SCHPO	Q99673 schizosacch
22	494	31.1	286	1 PANC_BACSU	P52398 bacillus su
23	491	30.9	282	1 PANC_SINY3	Q55073 synechocyst
24	489.5	30.8	278	1 PANC_NEIMB	P57036 neisseria m
25	488	30.7	281	1 PANC_XYLFA	Q9pgr8 xylella fas
26	487	30.7	283	1 PANC_PSEAE	Q9hv69 pseudomonas
27	483	30.4	301	1 PANC_VIBPA	Q87lv1 vibrio para
28	482.5	30.4	286	1 PANC_STAEP	Q8cr21 staphylococ
29	480.5	30.3	278	1 PANC_NEIMA	P57035 neisseria m
30	478	30.1	301	1 PANC_VIBVU	Q8dcl2 vibrio vuln
31	475	29.9	291	1 PANC_RHIME	Q92nn0 rhizobium m
32	472.5	29.8	280	1 PANC_XANAC	Q8pl10 xanthomonas
33	467.5	29.4	283	1 PANC_RHILO	Q98nd0 rhizobium 1

34	467.5	29.4	284	1 PANC_XANCP	Q8p9s9 xanthomonas
35	465	29.3	287	1 PANC_COREF	Q8fua6 corynebacte
36	465	29.3	292	1 PANC_AGRTS	Q8ua92 agrobacteri
37	457.5	28.8	285	1 PANC_EUCAI	P57292 buchnera ap
38	456	28.7	281	1 PANC_DEIRA	Q9rv66 deinococcus
39	455.5	28.7	279	1 PANC_CORGL	Q9x7l3 corynebacte
40	455.5	28.7	283	1 PANC_OCEIH	Q8cx59 oceanobacil
41	452.5	28.5	283	1 PANC_FALSO	Q8xwt3 ralstonia s
42	451	28.4	345	1 PANC_YEAST	P40459 saccharomyc
43	447	28.1	293	1 PANC_BRUME	Q8yfc9 brucella me
44	444	28.0	293	1 PANC_BRUSU	Q8g2j0 brucella su
45	423.5	26.7	313	1 PANC_MYCLE	O69524 mycobacteri

ALIGNMENTS

RESULT 1

PANC_LOTJA					
ID	PANC_LOTJA	STANDARD;	PRT;	307 AA.	
AC	O24035;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Pantoate--beta-alanine ligase precursor (EC 6.3.2.1) (Pantothenate synthetase) (Pantoate activating enzyme).				
GN	PANC.				
OS	Lotus japonicus.				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.				
OX	NCBI_TaxID=34305;				
RN	[1]				
RP	SEQUENCE FROM N.A., SEQUENCE OF 1-16, AND TISSUE SPECIFICITY.				
RC	STRAIN=cv. Gifu / B-129; TISSUE=Root nodules;				
RX	MEDLINE=99348031; PubMed=10417331;				
RA	Genschel U., Powell C.A., Abell C., Smith A.G.;				
RT	"The final step of pantothenate biosynthesis in higher plants: cloning and characterization of pantothenate synthetase from Lotus japonicus and Oryza sativum (rice).";				
RL	Biochem. J. 341:669-678(1999).				
CC	-!- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP + diphosphate + (R)-pantothenate.				
CC	-!- PATHWAY: Pantothenate biosynthesis; last step.				
CC	-!- SUBUNIT: Homodimer (Potential).				
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).				
CC	-!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN LEAF AND ROOT.				
CC	-!- MISCELLANEOUS: OPTIMAL PH IS 7.8. ACTIVITY DECREASES SHARPLY WITH INCREASING ACIDITY AND IS NIL AT PH 7. THERE IS ONLY A SLIGHT DECREASE TOWARDS HIGHER PH.				
CC	-!- SIMILARITY: Belongs to the pantothenate synthetase family.				

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC	EMBL; Y10252; CAA71302.1; --
DR	InterPro; IPR003721; Pantoate ligase.
DR	Pfam; PF02569; Pantoate_ligase; 1.
DR	TIGRFAMs; TIGR00018; panC; 1.
KW	Pantothenate biosynthesis; Ligase.
FT	INIT MET 0 0
FT	PROPEP 1 1 REMOVED PARTIALLY.
FT	CHAIN 2 307 PANTOATE--BETA-ALANINE LIGASE.
SQ	SEQUENCE 307 AA; 34109 MW; 91C464EBE4E1642 CRC64;

Query Match 81.0%; Score 1286; DB 1; Length 307;  
Best Local Similarity 80.8%; Pred. No. 7.2e-99;  
Matches 248; Conservative 24; Mismatches 31; Indels 4; Gaps 2;











RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum.";  
RL J. Bacteriol. 183:4823-4838(2001).  
CC -!- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +  
CC diphosphate + (R)-pantothenate.  
CC -!- PATHWAY: Pantothenate biosynthesis; last step.  
CC -!- SIMILARITY: Belongs to the pantothenate synthetase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE007789; AAK80857.1; -.  
DR PIR; F97258; F97258.  
DR HAMAP; MF\_00158; -; 1.  
DR InterPro; IPR003721; Pantoate\_ligase.  
DR Pfam; PF02569; Pantoate\_ligase; 1.  
DR TIGRFAMS; TIGR00018; panC; 1.  
KW Pantothenate biosynthesis; Ligase; Complete proteome.  
SQ SEQUENCE 281 AA; 31543 MW; C42FC71AC2070409 CRC64;  
  
Query Match 33.8%; Score 537; DB 1; Length 281;  
Best Local Similarity 40.2%; Pred. No. 4.2e-37;  
Matches 121; Conservative 51; Mismatches 101; Indels 28; Gaps 5;  
  
QY 8 ISD-KASMRWSRSMRAQKLI GLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQG 66  
Db ||| : : : : ||||| : : : : ||| : : : : ||| : : : : |||  
7 ISDVKKYIKQWK- - - EGLTIGLVPTMGYLHDGKSLTERASKENDKIVISDFVNP IQF 62  
  
QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLYDGKNGGDDVAEAGGMVSCVESG 126  
Db ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
63 GPNEDLDVYPRDLDRDAE- - VCTKAGASILFNPEPSEMVFDDA- - - - - 103  
  
QY 127 SGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLIQRM 186  
Db : : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
104 - - - VTFVNSSKITDILCGARRPGHFRGVCVTVTKLFNITCPDRAYFGEKDAQQVAVIKRM 160  
  
QY 187 VRDLDFSIKVIGAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRKSAAGDQGVHC 246  
Db ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
161 VRDLNFDIEIVACPIIREEDGLAKSSRNLYLSSEERKAATILSKSLNLAKELLDNGEKNV 220  
  
QY 247 EKLTN-LVIQSVTDAGGRIDYAEIVDQNNLEKVEIQKSPVVFCAAWFGKVRLLDNMEIN 305  
Db : : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
221 YNICKAIIIEIGKEPLAKIDYVEWVDSLSLKSVSQVQSILVAIVYIGKIRLIDNFTWN 280  
  
QY 306 L 306  
Db :  
281 I 281  
  
RESULT 8  
PANC\_CAMJE  
ID PANC CAMJE STANDARD; PRT; 282 AA.  
AC Q9PIK2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Pantoate--beta-alanine ligase (EC 6.3.2.1) (Pantothenate synthetase)  
DE (Pantoate activating enzyme).  
GN PANC OR Cj0297C.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteraceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,

RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
CC -!- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +  
CC diphosphate + (R)-pantothenate.  
CC -!- PATHWAY: Pantothenate biosynthesis; last step.  
CC -!- SIMILARITY: Belongs to the pantothenate synthetase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL139074; CAB72764.1; -.  
DR PIR; E81448; E81448.  
DR HAMAP; MF\_00158; -; 1.  
DR InterPro; IPR003721; Pantoate\_ligase.  
DR Pfam; PF02569; Pantoate\_ligase; 1.  
DR TIGRFAMS; TIGR00018; panC; 1.  
KW Pantothenate biosynthesis; Ligase; Complete proteome.  
SQ SEQUENCE 282 AA; 32087 MW; C8946F933D7B6475 CRC64;  
  
Query Match 33.8%; Score 536; DB 1; Length 282;  
Best Local Similarity 41.5%; Pred. No. 5.1e-37;  
Matches 125; Conservative 46; Mismatches 98; Indels 32; Gaps 5;  
  
QY 6 RVISDKASMRWSRSMRAQKLI GLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQG 65  
Db : : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
2 QVITSVKEAQIVKDWKSHQLSIGYVPTMGFLHDGHLVKKHAK-TQDKVIVSIFVNPQM 60  
  
QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLYDGKNGGDDVAEAGGMVSCVES 125  
Db ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
61 FGPNEDESSYPRDLERDIKMCQD- - NGVDMVFIPDATQMYLKN- - - - - 101  
  
QY 126 GSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLIQ 185  
Db : : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
102 - - - FSTYVDMNTITDKLGGAKRPGHFRGVCVTVTKLFFNINLPDIVMGQKDAQQCVVVRH 158  
  
QY 186 MVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRKSAAGDQGVH 245  
Db ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
159 MVDDLNFDLKIQICPIIREEDGLAKSSRNVLHSPEREKALSINKSLLRKSAAGDQGVH 245  
  
QY 246 CEKLTNLVVIQSVTDAGG- - - - -RIDYAEIVDQNNLEKVEIQKSPVVFCAAWFGKVRLLID 300  
Db ||| : : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||  
216 -EKNTSKIIQAMKDILEKEKLIKIDYIELVDFTMENIENTDNVLGAVAAAFVGKTRLID 274  
  
QY 301 N 301  
Db :  
275 N 275  
  
RESULT 9  
PANC\_AQUAE  
ID PANC AQUAE STANDARD; PRT; 282 AA.  
AC O67831;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Pantoate--beta-alanine ligase (EC 6.3.2.1) (Pantothenate synthetase)  
DE (Pantoate activating enzyme).  
GN PANC OR AQ\_2132.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.

```
RC STRAIN=VF5;
RA MEDLINE=98196666; PubMed=9537320;
RX Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL aeolicus.";
RL Nature 392:353-358(1998).
CC -!- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +
CC diphosphate + (R)-pantothenate.
CC -!- PATHWAY: Pantothenate biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the pantothenate synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000774; AAC07847.1; -.
CC PIR; G70482; G70482.
CC HAMAP; MF 00158; -.
CC InterPro; IPR003721; Pantoate_ligase.
CC Pfam; PF02569; Pantoate_ligase; 1.
CC TIGRFAMs; TIGR00018; panC; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 282 AA; 32433 MW; 7E71B941C158DCEB CRC64;

Query Match 32.7%; Score 518.5; DB 1; Length 282;
Best Local Similarity 40.6%; Pred. No. 1.4e-35;
Matches 123; Conservative 47; Mismatches 106; Indels 27; Gaps 5;

QY 5 PRVSDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP 64
Db 2 PLLFKKIKDLNLFKNKRCCEGKEIGFVPTMGYLGHEGRQLIKLARMQNDIVVSIYVNP 61
QY 65 QFAPTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLYDYGNKGGDVAEAGGMVSCVE 124
Db 62 QFGEGEDVDYPRDLERDLE--ICEEGVVDVVFAPVEVDIYFK----- 102
QY 125 SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKOYQQWRLLIQ 184
Db 103 ---GYRTKVCVGELGKVLGEFRPGHFDGVATIVVKLFNIVQPNRAYFGEKDYQQLKIE 159
QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLRAKSAAGDQV 244
Db 160 QVVEDLNIPVEIVPPIVREEDGLAYSSRNVLHLSPEEREKALSINKSLRAKSAAGDQV 219
QY 245 HCEKLTNLV---IQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVFCVAAWFGKVRLLDN 301
Db 220 DAKRIKEAIRAFIERHPHVKG-VDYVEITDQ-NLNPKETVEKGRILVAVRVGNARLLDN 277
QY 302 MEI 304
Db 278 WKV 280

RESULT 10
PANC_ECOLI
ID PANC_ECOLI
AC P31663;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Pantoate--beta-alanine ligase (EC 6.3.2.1) (Pantothenate synthetase)
DE (Pantoate activating enzyme).
GN PANC OR B0133.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Merkel W.K., Nichols B.P.;
RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +
CC diphosphate + (R)-pantothenate.
CC -!- PATHWAY: Pantothenate biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the pantothenate synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L17086; AAA24272.1; -.
CC PIR; AE000122; AAC73244.1; -.
CC PDB; E64736; E64736.
CC PDB; 1IHO; 30-MAY-01.
CC SWISS-2DPAGE; P31663; COLI.
CC EcoGene; EG11746; panC.
CC HAMAP; MF 00158; -.
CC InterPro; IPR004821; Cyt tran rel.
CC InterPro; IPR003721; Pantoate_ligase.
CC Pfam; PF02569; Pantoate_ligase; 1.
CC TIGRFAMs; TIGR00125; cyt tran_rel; 1.
CC TIGRFAMs; TIGR00018; panC; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome; 3D-structure.
SQ SEQUENCE 283 AA; 31597 MW; 4246647C6FDAFB20 CRC64;

Query Match 32.5%; Score 516; DB 1; Length 283;
Best Local Similarity 38.9%; Pred. No. 2.3e-35;
Matches 118; Conservative 52; Mismatches 109; Indels 24; Gaps 3;

QY 7 VISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP 66
Db 3 ILETPLLRQQIRRLRMEGKRVAVPTMGNLHGHMKLVDEAKARADVIVVSIYVNP 62
QY 67 APTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLYDYGNKGGDVAEAGGMVSCVE 124
Db 63 DRPEDLARYPRTLQEDCEKLNK--RKVDLVFAPSVKIY-----P 100
QY 125 SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLLIQ 184
Db 101 NGTETHYVDVPGVGLSTMLEGASRPGHFRGVSTIVSKLFNLVQPDIIACFGEKDFQQLALIR 160
QY 185 RMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLRAKSAAGDQV 244
Db 161 KMWADMGFDIEIVGVPIMRKADGLALSSRNGLYTAEQRIAPGLYKVLSSIADKLQAGER 220
QY 245 HCEKLTNLVSIQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVFCVAAWFGKVRLLDNMEI 304
Db 221 DLDEIITIAGQELNEKGFRADDIIQIRDADTLLLEVSETSKRAVILVAANLGDARLLDNKMW 280
QY 305 NLS 307
Db 281 ELA 283
```







```
DR HAMAP; MF 00158; -; 1.
DR InterPro; IPR004821; Cyt_tran_rel.
DR InterPro; IPR003721; Pantoate_ligase.
DR Pfam; PF02569; Pantoate_ligase; 1.
DR TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
DR TIGRFAMs; TIGR00018; panC; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 281 AA; 30435 MW; CC7526CCB236BD3C CRC64;

Query Match      31.9%; Score 506.5; DB 1; Length 281;
Best Local Similarity 39.0%; Pred. No. 1.4e-34;
Matches 114; Conservative 48; Mismatches 109; Indels 21; Gaps 3;

QY 14 MRSWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNPQGAPTEDLS 73
Db 10 IRTQVRAWRAKETVAFVPTMGNLHQGHITLVKEAKKCDHVVASIFVNPQMFGQNEDDL 69

QY 74 TYPSEDFDGVKKLASVPGGVDDVVFHPRNLYDYGKNGGDDVAEAGGMVSCVESGSHESW 133
Db 70 AYPRTLEADSQALTA--AGAELEFTPTPALIYPK-----GLAQQTYV 109

QY 134 RVEKLEGLCGKSRPVEFRGVATVVTKLENIVEPDVAVFGKDYQWRLIQRMVRDLDFS 193
Db 110 EVPGISDVLCCASRPGHFRGVATVCKLENIVQPDIAFFGNKDYQQLLVIRTMVEDLSLP 169

QY 194 IKVIGAEITRDNDGLAMSSRNHLSPEEREKALSINKSLLRKSAAGDGQVHCEKLTNLV 253
Db 170 IEIIGIDITREASGLAMSSRNGYLTAQEKAAAPALKKAIDAMAQGIKQG-ISIEQVTEEA 228

QY 254 IQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVWFCVAAWFGKVKRLIDNMEIN 305
Db 229 KASLTAAGTPTDYLEVRHADTLAKAETQDKALVILAAAYLGKARLIDNLRFD 280

RESULT 15
PANC_STAAM STANDARD; PRT; 283 AA.
AC Q99R39;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pantoate--beta-alanine ligase (EC 6.3.2.1) (Pantothenate synthetase)
DE (Pantoate activating enzyme).
GN PANC OR SAV2598 OR SA2391.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus."
RL Lancet 357:1225-1240(2001).
CC -!- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +
CC diphosphate + (R)-pantothenate.
CC -!- PATHWAY: Pantothenate biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the pantothenate synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003365; BAB58760.1; -.
DR EMBL; AP003137; BAB43696.1; -.
DR PIR; F90066; F90066.
DR HAMAP; MF 00158; -; 1.
DR InterPro; IPR003721; Pantoate_ligase.
DR Pfam; PF02569; Pantoate_ligase; 1.
DR TIGRFAMs; TIGR00018; panC; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 283 AA; 31430 MW; 976C8CB0D4F90F38 CRC64;

Query Match      31.8%; Score 504.5; DB 1; Length 283;
Best Local Similarity 37.2%; Pred. No. 2e-34;
Matches 113; Conservative 58; Mismatches 102; Indels 31; Gaps 5;

QY 6 RVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNPQG 65
Db 3 KLITTVKEMQHIVKAAKRSGITIGFIPMTGALHDGHLTMVRESVSTNDITVVSFVNPLQ 62

QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVDDVVFHPRNLYDYGKNGGDDVAEAGGMVSCVES 125
Db 63 FGNEFDAYPRQIDKOLELVSEV--GADIVFHPAVEDIYPGELGIDV-KVGPLADVLE- 118

QY 126 GSGHESWVRVEKLELGLCGKSRPVEFRGVATVVTKLENIVEPDVAVFGKDYQWRLIQ 185
Db 119 -----GAKRPGHFDGVVTVVVKLFNIVMPDYAYFGKDKAQQLAIVEQ 160

QY 186 MVRDLDFSIVKIGAEITRDNDGLAMSSRNHLSPEEREKALSINKSLLRKSAAGDGQVH 245
Db 161 MVKDPNHAVEIIGIDIVREADGLAKSSRNLYLTEQERQEAHVLSKSLLLAQAALYQDG--- 217

QY 246 CEKLTNLVIQSVTD-----AGGRIDYAEIVDQNNLEKVEIKSPVWFCVAAWFGKVKRLID 300
Db 218 -ERQSKVIIDRVTEYLESHISGRIEEAVVSYYPQLVEQHEITGRIFISLAVKFSKARLID 276

QY 301 NMEI 304
Db 277 NIII 280

Search completed: December 15, 2003, 13:34:47
Job time : 18 secs
```





Db 3 PEVIRKDSMRKWSRAMRSQKKTIGLVPTMGYLHEGHLSLVRQSLALTDVTVVSIYNPG 62

QY 65 QFAPTEDLSTYPSDFDGVKKLASVPGGVDVVFHPRNLVDYG-----KNGGGDVAEAGG 118

Db 63 QFSPTEDLSTYPSDFSGDLTKLAALSGGKVVVFENPKNLDYGGGETKKINDGGN---GGR 119

QY 119 MVSCVESGS-GHESWVRVEKLELGLCGKSRPVFRFGVATVVTKLFNIVEPDVAVFGKKDY 177

Db 120 VVSCVEEGLGHETWIRVERLEKGLCGKSRPVFRFGVATIVTKLFNIVEPDVALFGKKDY 179

QY 178 QQWRLIORMVRDLDFSIVKIVIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLLRAS 237

Db 180 QQWRIIORMVRDLNFGIEIVGSDIAREKDGGLAMSSRNVRLSDEERQALSISRSLAMAKA 239

QY 238 AAGDGOVHCEKLTNLVIQSVTDAGGRIDYAEIIVDQNNLEKVEQIKSPVVFCAAWFGKVR 297

Db 240 SVAEGKTNCAELKDMIIQQVVGSGRVDYVEIVDQETLEGVEEIKSGVVVICVAAWFGTVR 299

QY 298 LIDNMEINLSM 308

Db 300 LIDNIEINVS 310

RESULT 2

Q8GDP0

ID Q8GDP0 PRELIMINARY; PRT; 315 AA.

AC Q8GDP0;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Pantoate-beta-alanine ligase (EC 6.3.2.1) (Fragment).

OS Helicobacillus mobilis.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae;

OC Helicobacillus.

OX NCBI\_TaxID=28064;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22337798; PubMed=12446909;

RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,

RA Blankenship R.E.;

RT "Whole-genome analysis of photosynthetic prokaryotes.";

RL Science 298:1616-1620(2002).

RN [2]

RP SEQUENCE FROM N.A.

RA Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,

RA Gerdes S., Kyrpides N., Overbeek R.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY142936; AAN87540.1; --

KW Ligase.

FT NON TER 315 315

SQ SEQUENCE 315 AA; 35339 MW; DBA58CC255F8F40C CRC64;

Query Match 38.8%; Score 616.5; DB 2; Length 315;

Best Local Similarity 42.6%; Pred. No. 9.6e-44;

Matches 132; Conservative 57; Mismatches 94; Indels 27; Gaps 5;

QY 3 PAPRVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVN 62

Db 25 PYMRLISSVNMMAWAKEQRRVGHGTIGLVPTMGYLHEGHLTLMRRAKENCEKVVVSIFVN 84

QY 63 PGQFAPTEDLSTYPSDFDGDVKKLASVPGGVDVVFHP--RNLYDYGKNGGGDVAEAGGMV 120

Db 85 PLQFAGAGEDYEEYPRDLTRD-SQLAD-SAGVDVIFAPAVKDMYP----- 126

QY 121 SCVESGSHESWVRVEKLELGLCGKSRPVFRFGVATVVTKLFNIVEPDVAVFGKKDYQQW 180

Db 127 -----KGYSSFVEVEQVSDHLCGAARPGHFRGVTTVVSKLFNIVRPDIAFFGQKDAQQL 180

QY 181 RLIIORMVRDLDFSIVKIVIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLLRASAA 240

Db 181 AIIRRMVEDLNMGIAIVGPPIVREADGLALSSRNVYLSPEERKAALVLSQSLKKAKELIA 240

QY 241 DGQVHCEKLTNLVIQSV-TDAGGRIDYAEIIVDQNNLEKVEQIKSPVVFCAAWFGKVRLI 299

Db 241 EGEAEERIRQEIIVKIVIAEPLANIDYVQIVDNRFIQIPVERLEGECLIALAVRFGKTRLI 300

QY 300 DNMEINLSMN 309

Db 301 DNLVMEVSPN 310

RESULT 3

Q9AMR9

ID Q9AMR9 PRELIMINARY; PRT; 283 AA.

AC Q9AMR9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE ID912.

GN ID912.

OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae; Bradyrhizobium.

OX NCBI\_TaxID=375;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=110spc4;

RX MEDLINE=21101824; PubMed=11157954;

RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,

RA Hennecke H.;

RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb

RT DNA region of the Bradyrhizobium japonicum chromosome.";

RL J. Bacteriol. 183:1405-1412(2001).

DR EMBL; AF322013; AAG61078.1; --

DR InterPro; IPR004821; Cyt\_tran\_rel.

DR InterPro; IPR003721; Pantoate\_ligase.

DR Pfam; PF02569; Pantoate\_ligase; 1.

DR TIGRFAMS; TIGR00125; cyt\_tran\_rel; 1.

DR TIGRFAMS; TIGR00018; panC; 1.

SQ SEQUENCE 283 AA; 31322 MW; A2CA526734D9A607 CRC64;

Query Match 34.3%; Score 545; DB 2; Length 283;

Best Local Similarity 41.0%; Pred. No. 8.8e-38;

Matches 125; Conservative 48; Mismatches 104; Indels 28; Gaps 5;

QY 6 RVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNPQ 65

Db 2 KVITKVAELRRALADVRNAEKRIQFVPTMGYLHDGHLALISASREHCDVTVVSIFVNPTQ 61

QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVDVVFHP--RNLYDYGKNGGGDVAEAGGMVSCV 123

Db 62 FGPNEILSRYPRDFARDEALCGS--AGVSIIFAPSAEEIYP----- 100

QY 124 ESGSGHESWVRVEKLELGLCGKSRPVFRFGVATVVTKLFNIVEPDVAVFGKKDYQQWRLI 183

Db 101 ---AQFESFVEPEGELAKPLCGAFRPGHFRGVATVVCLEFNMVQPDVAYFGQKDFQCCAVI 157

QY 184 QRMVRDLDFSIVKIVIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLLRASAAAGDQ 243

Db 158 RRMVTVDLNLPIEIVTPTVREPDGLAMSSRNRYLCPEERDRSLAISRGLFAAAHEFASGE 217

QY 244 VHCEKLTNLVIQSVTDAGGRIDYAEIIVDQNNLEKVEQ-IKSPVVFCAAWFGKVRLIDNM 302

Db 218 RDAATLIALARRHL-ERVDRLOQYLELVDPGTLRIADSPLRYPAVLCVAAAYVGSTRLIDNV 276

QY 303 EINLS 307

Db 277 VLSWS 281

RESULT 4

Q8EIH0

ID Q8EIH0 PRELIMINARY; PRT; 281 AA.

AC Q8EIH0;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)







Db 1 MPTMGNLHDGHMKLVDEAKARADVVSIFVNPQMQRDPEDLARYPRTLQEDCEKLNK-- 58

QY 91 GGVVVVFP--RNLYDYKNGGDDVAEAGMVSCVESGSHESWVRVEKLELGLCGKSRP 148

Db 59 RKVDLVFAPSVEIY-----PNGTETHYVDVPGLSLTMLEGASRP 98

QY 149 VFERGVATVVTKLFNIVEPDVAVFGKKDYQQWRLIQRMVRDLDFSIKVIGAEITRDNDGL 208

Db 99 GHFRGVSTIVSKLNLVQPDIAACFGEKDFQALIRKMWADMGFDIEIVGVPIMRAKDGL 158

QY 209 AMSSRNVLHSPEREKALSINKSLLRAKSAAGDGOVHCEKLTNLVTSQVTDAGGRIDYAE 268

Db 159 ALSSRNGYLTAEQRKIAPGLYKVLSSIAADKLQAGERDLDEIIAIAQELNEKGFRADDIQ 218

QY 269 IVDQNNLEKVEIKSPVVFCAAWFGKVRLLIDNMEINLS 307

Db 219 IRDADTLLEVSSENSKRAVILVAAWLGDARLLIDNKMVELA 257

RESULT 10

Q8YSZ3 PRELIMINARY; PRT; 534 AA.

AC Q8YSZ3

DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Pantothenate synthetase.

GN PANC OR ALR2936.

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

OX NCBI\_TaxID=103690;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21595285; PubMed=11759840;

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,

RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,

RA Yasuda M., Tabata S.;

RT "Complete genomic sequence of the filamentous nitrogen-fixing

RT cyanobacterium Anabaena sp. strain PCC 7120.";

RL DNA Res. 8:205-213 (2001).

DR EMBL; AP003591; BAB74635.1; --

DR InterPro; IPR003136; Cytidylylate\_kin.

DR InterPro; IPR004821; Cyt\_tran\_rel.

DR InterPro; IPR003721; Pantoate\_ligase.

DR Pfam; PF02224; Cytidylylate\_kin; 1.

DR Pfam; PF02569; Pantoate\_ligase; 1.

DR TIGRFAMS; TIGR00017; cmk; 1.

DR TIGRFAMS; TIGR00125; cyt\_tran\_rel; 1.

DR TIGRFAMS; TIGR00018; panC; 1.

KW Complete proteome.

SQ SEQUENCE 534 AA; 59176 MW; 23434C86E3DFF1BC CRC64;

Query Match 29.4%; Score 467; DB 16; Length 534;

Best Local Similarity 36.8%; Pred. No. 8.5e-31;

Matches 110; Conservative 54; Mismatches 95; Indels 40; Gaps 7;

QY 13 SMRSRSRMRAQKLIQLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGQFAPTEDL 72

Db 33 SMTSWYQT-----AIGLVPTMGSLHQHLSLIERARHENSTIVSIFINPLQFGPNEDY 86

QY 73 STYPSDFDGDVKKLASVPGGVVVFHP-----RNLYDYKNGGDDVAEAGMVSCV 123

Db 87 GRYPRLEQD-RQLCE-QGGVDALFAPSPEELGIPQKNIQE---SQVTQVIPPSVMIS-- 139

QY 124 ESGSHESWVRVEKLELGLCGKSRPVFERGVATVVTKLFNIVEPDVAVFGKKDYQQWRLI 183

Db 140 -----GLCGHSRLGHFGQVATIVTKLNLVQPDRAVFGQKDGQQLAVI 182

QY 184 QRMVRDLDFSIKVIGAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRAKSAAGDGO 243

Db 183 KRLVADLDLPVEIVACPTVREASGLACSSRNQYLTAQEQQAALYRGLLQAEAFKAGV 242

QY 244 VHCEKLTNLVTSQVTDAGG-RIDYAEIVDQNNLEKVEIKSPVVFCAAWFGKVRLLIDN 301

Db 243 RYSSRLREVVRQELAKVSSVLVEYIELVEPTTLMPLDKIQEGMGLAIAARLGSTRLLIDN 301

RESULT 11

Q8FUA6 PRELIMINARY; PRT; 303 AA.

AC Q8FUA6;

DT 01-MAR-2003 (TReMBLrel. 23, Created)

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Pantoate-beta-alanine ligase (EC 6.3.2.1).

GN PANC OR CE0115.

OS Corynebacterium efficiens.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.

OX NCBI\_TaxID=152794;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;

RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,

RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,

RA Usuda Y., Sugimoto S.;

RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP005214; BAC16925.1; --

KW Ligase; Complete proteome.

SQ SEQUENCE 303 AA; 33177 MW; CFF657FCD718C950 CRC64;

Query Match 29.3%; Score 465; DB 16; Length 303;

Best Local Similarity 42.1%; Pred. No. 5.5e-31;

Matches 122; Conservative 33; Mismatches 107; Indels 28; Gaps 6;

QY 21 MRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGF---APTEDLSTYPS 77

Db 33 LHTQORSIGLVPTMGALHSGHASLVAKARENDVVVTSIFVNPLOQFEALGDCDDYRNYR 92

QY 78 DFDGDVKKLASVPGGVVDVVFHPRNLYDYKNGGDDVAEAGMVSCVESGSHESWVRVEK 137

Db 93 QLDADVALLEA--EGVDVVFAP-----DVEMYGPTGL-----FMLWVRTGE 132

QY 138 LELGLCGKSRPVFERGVATVVTKLFNIVEPDVAVFGKKDYQQWRLIQRMVRDLDFSIKVI 197

Db 133 MGERLEGASRPCHFDGVATVVAKLFNLRPDRAYFGQKDAQQVAVIRRLVRDLDFLEIR 192

QY 198 GAEITRDNDGLAMSSRNVLHSPEREKALSINKSLLRAKSAAGDGOVHCEKLTNLVTSQV 257

Db 193 AVPIIRAADGLAESSRNQRLSGADRAAALVLPRLVFLDEERAATGQP--LDIEGARARQLR 250

QY 258 TDAGGRIDYAEIVDQNNLEKVE---QIKSPVVFCAAWFGKVRLLIDNMEI 304

Db 251 AAQGVRLDHLVDAATLEPLTVTGALDRPALVVAIHHVGVQVRLLIDNIEL 300

RESULT 12

Q8CX59 PRELIMINARY; PRT; 283 AA.

AC Q8CX59;

DT 01-MAR-2003 (TReMBLrel. 23, Created)

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Pantoate beta-alanine ligase (Pantothenate synthetase)

DE (EC 6.3.2.1).

GN OB3275.

OS Oceanobacillus iheyensis.

OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.

OX NCBI\_TaxID=182710;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HTE831 / DSM 14371 / JCM 11309;

RX MEDLINE=22220767; PubMed=12235376;







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 13:32:29 ; Search time 41 Seconds  
(without alignments)  
1200.128 Million cell updates/sec

Title: US-10-033-269-9  
Perfect score: 1588  
Sequence: 1 MAPAPRVISDKASMRWSRS.....AWFGKVLIDNMEINLSMNV 310

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1286	81.0	308	20	AA1980
2	1101	69.3	310	23	AA1981
3	1101	69.3	643	21	AA1982
4	1101	69.3	653	21	AA1983
5	1101	69.3	654	21	AA1984
6	1098	69.1	310	21	AA1985
7	1098	69.1	322	21	AA1986
8	1071	67.4	299	21	AA1987
9	927.5	58.4	333	20	AA1988

10	494	31.1	286	22	AAU01245
11	482.5	30.4	288	23	ABP39476
12	455.5	28.7	279	21	AA10683
13	455.5	28.7	279	21	AA10032
14	455.5	28.7	279	21	AA10034
15	455.5	28.7	279	22	AA10034
16	455.5	28.7	279	22	AA10034
17	454.5	28.6	259	24	ABP78440
18	427	26.9	278	24	AA10034
19	418	26.3	285	23	AB10034
20	405	25.5	275	23	AB10034
21	372.5	23.5	229	22	AA10034
22	357	22.5	481	22	AA10034
23	332	20.9	903	22	AA10034
24	332	20.9	939	22	AA10034
25	303.5	19.1	98	21	AA10034
26	230	14.5	111	23	ABU51871
27	228	14.4	266	22	AA10034
28	228	14.4	266	22	AA10034
29	160.5	10.1	117	23	ABU50660
30	103.5	6.5	421	18	AA19735
31	103.5	6.5	421	20	AA19938
32	97.5	6.1	1297	23	AA196927
33	97.5	6.1	1298	23	AA196928
34	94.5	6.0	1297	23	AA196930
35	94	5.9	659	24	ABJ26131
36	94	5.9	975	20	AA142711
37	93.5	5.9	423	24	ABJ18791
38	93	5.9	422	22	AA182216
39	92	5.8	332	23	ABP27548
40	91.5	5.8	657	22	AA131467
41	91	5.7	339	22	AA190770
42	91	5.7	339	22	AA176545
43	90.5	5.7	657	13	AA129580
44	90.5	5.7	997	22	ABG29047
45	90.5	5.7	1118	22	ABG24324

ALIGNMENTS

RESULT 1  
AAY42402  
ID AAY42402 standard; Protein; 308 AA.  
XX  
AC AAY42402;  
XX  
DT 02-DEC-1999 (first entry)  
XX  
DE Amino acid sequence of L. japonicus Pantothenate Synthetase (PS).  
XX  
KW pantothenate synthetase; cloning; herbicide; biosynthesis;  
KW Coenzyme A precursor; pantoate-beta-alanine ligase.  
XX  
OS Lotus japonicus.  
XX  
PN WO9942565-A1.  
XX  
PD 26-AUG-1999.  
XX  
PF 02-JUN-1998; 98WO-EP03261.  
XX  
PR 31-MAY-1997; 97GB-0011163.  
PR 27-JUN-1997; 97GB-0013477.  
XX  
(AGRE ) HOECHST-SCHERING AGREVO GMBH.  
PA Abell C, Smith AG, Genschel U, Laber B;  
XX WPI; 1999-527466/44.  
DR N-PSDB; AAZ20926.  
XX  
PT New isolated DNA molecule, useful in herbicidal compositions -

XX PS Claim 3; Fig 1.2; 64pp; English.

XX CC This is the amino acid sequence of the Lotus japonicus Pantothenate

CC synthetase (PS) enzyme, which is essential in the biosynthesis of

CC Coenzyme A.

CC The invention provides methods to assay proteins for PS activity, and

CC the ability of compounds to inhibit PS.

CC The assay compounds which are found to inhibit PS can be used as

CC herbicides.

CC The nucleotide probe can be used to amplify PS coding sequences from a

CC chosen organism via the process of polymerase chain reaction.

CC A number of assays have previously been used to measure PS activity

CC but are unsuitable for large screen biochemical screening of compounds

CC to find PS inhibitors. In this invention however the isolation of the

CC DNA molecule encoding PS enables a host cell to produce the protein

CC which can then be tested against compounds potentially able to inhibit

CC the enzyme.

XX SQ Sequence 308 AA;

Query Match 81.0%; Score 1286; DB 20; Length 308;

Best Local Similarity 80.8%; Pred. No. 3.8e-126;

Matches 248; Conservative 24; Mismatches 31; Indels 4; Gaps 2;

QY 4 APRVISDKASMRSMRSMRAQKLGIVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP 63

Db 2 APNVISDKEMRKWSRSMRSQKLIALVPTMGFLHAGHLSLVRDAHNEADLVAVSIYVNP 61

QY 64 GQFAPTEDLSTYPSDFDGDVKKLASVPGVDVVFHPRNLYDYGKNGGDAEAG--GMVS 121

Db 62 GQFSPTEDLSTYPSDFDGDVKKLASVPGVDVVFHPRNLYDYGKNGGDAEAGCGDGVVS 121

QY 122 CVE--SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVTKLFNIVEPDVAVFGKKDYQQ 179

Db 122 CVDRRSFGGHETWVRAEKLKPLCGKSRPVFFRGVATVTKLFNIVEPDVAVFGKKDYQQ 181

QY 180 WRLIQRMVRDLDFSITKIVIGAEITRDNDGLAMSSRNVLSPEREKALSINKSLRAKSA 239

Db 182 WKIIQRMVRDLDFSITKIVIGAEITRDNDGLAMSSRNVLSPEREKALSINKSLFRAKSA 241

QY 240 GDQVHCEKLTNLVQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVFCVAAWFGKVR 299

Db 242 EDGQIHCEKLTNLVQSVITEAGGRIDYAEIVDQNNLEKVEIKSPVFCVSAWFGKRLI 301

QY 300 DNMEINL 306

Db 302 DNIEINL 308

RESULT 2

ABB93747

ID ABB93747 standard; Protein; 310 AA.

XX ABB93747;

AC ABB93747;

XX 31-MAY-2002 (first entry)

DT 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 2958.

DE Herbicidally active polypeptide SEQ ID NO 2958.

XX Herbicidal; plant; agriculture; herbicide.

KW Arabidopsis thaliana.

XX Arabidopsis thaliana.

OS WO200210210-A2.

XX WO200210210-A2.

PN 07-FEB-2002.

XX 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

FA (FARB ) BAYER AG.

XX PI Tietjen K, Weidler M;

XX DR WPI; 2002-269010/31.

XX PT Identifying plant target proteins for herbicidally active compounds,

PT comprising aligning and comparing nucleic acid or amino acid sequences

PT from plant with nucleic acid or amino acid sequences from non-plant

PT organisms -

XX Claim 5; SEQ ID NO 2958; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value

CC greater by a factor of 3 than the E-value of most similar non-plant

CC sequences are selected. The polypeptides or nucleic acids encoding them

CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

XX SQ Sequence 310 AA;

Query Match 69.3%; Score 1101; DB 23; Length 310;

Best Local Similarity 67.2%; Pred. No. 1e-106;

Matches 209; Conservative 46; Mismatches 46; Indels 10; Gaps 3;

QY 5 PRVISDKASMRSMRSMRAQKLGIVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP 64

Db 3 PEVIRDKDSMRKWSRAMRSQKLTIGLVPTMGYLGHLHLVRSQSLALTDVTVSIYVNP 62

QY 65 QFAPTEDLSTYPSDFDGDVKKLASVPGVDVVFHPRNLYDYG----KNGGGDVAEAG 118

Db 63 QFSPTEDLSTYPSDFDGDVKKLASVPGVDVVFHPRNLYDYGKNGG--GGR 119

QY 119 MVSCVBSGS-GHESWVRVEKLELGLCGKSRPVFFRGVATVTKLFNIVEPDVAVFGKKDY 177

Db 120 VVSCVEEGGLGHETWIRVERLEKGFCKSRPVFFRGVATVTKLFNIVEPDVAVFGKKDY 179

QY 178 QQWRLIQRMVRDLDFSITKIVIGAEITRDNDGLAMSSRNVLSPEREKALSINKSLRAKS 237

Db 180 QQWRLIQRMVRDLNFGIEIVGSDIAREKDGAMSSRNVLSPEREKALSISRSLAMAKA 239

QY 238 AAGCQVHCEKLTNLVQSVTDAGGRIDYAEIVDQNNLEKVEIKSPVFCVAAWFGKVR 297

Db 240 SVAEGKTNCAELKDMIIQQVVSAGRDYVEIVDQETLEGVEEIKSGVVICVAAWFGTVR 299

QY 298 LIDNMEINLSM 308

Db 300 LIDNIEINVS 310

RESULT 3

AAG39461

ID AAG39461 standard; Protein; 643 AA.

XX AAG39461;

AC AAG39461;

XX 18-OCT-2000 (first entry)

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 48826.

DE Arabidopsis thaliana.

XX Arabidopsis thaliana.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.



PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 69.3%; Score 1101; DB 21; Length 643;		
Best Local Similarity 67.2%; Pred.No. 3.2e-106;		
Matches 209; Conservative 46; Mismatches 46; Indels 10; Gaps 3;		
QY	5	PRVISDKASMRWSRSMRAQKLI GLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNP
Db	336	PEVIRDKDSMRKWSRAMRSQKTI GLVPTMGYLGHEGLSLVRQSLALTDVTVWSIYVNP
QY	65	QFAPTEDLSTYPSDFDGDVKKLASVPGVDVVFHPRNLYDYG-----KNGGGDVAEAGG
Db	396	QFSPTEDLSTYPSDFSGDLTKLAALSGGKVVFVFNPKNLYDYGGETKKINDGGN---GGR
QY	119	MVSCVESGS-GHESWVRVEKLELGLCGKSRPVPFFRGVATVVTKLFNIVEPDVAVFGKKDY
Db	453	VVSCVEEGLGHETWIRVERLEKFGCGKSRPVPFFRGVATIVTKLFNIVEPDVALFGKKDY
QY	178	QQWRLTQRMVRDLDFSIXKIVIGAEITRDNDGLAMSSRNVLSPEREKALSINKSLRAKS
Db	513	QQWRIIQRMVRDLNFGIEIVGSDIAREKDGGLAMSSRNVLSDDEERQALSISRSLAMAKA
QY	238	AAGDQVHCEKLTNLVIOQSVTDAGGRIDYAEIVDQNNLEKVEIQKSPVFCVAAWFGKVR
Db	573	SVAEGXTNCAELKDMIIQQVVGSGRVDYVEIVDQETLEGVEEIKSGVVICVAAWFGTVR
QY	298	LIDNMEINLSM 308
Db	633	LIDNIEINVSL 643
RESULT 4		
ID	AAG39460	
XX	AAG39460 standard; Protein; 653 AA.	
AC	AAG39460;	
XX	18-OCT-2000 (first entry)	
DT		

XX	Arabidopsis thaliana protein fragment SEQ ID NO: 48825.
DE	Protein identification; signal transduction pathway; metabolic pathway;
XX	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
KW	Arabidopsis thaliana.
XX	EP1033405-A2.
PN	06-SEP-2000.
XX	25-FEB-2000; 2000EP-0301439.
PD	25-FEB-1999; 99US-0121825.
XX	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.

PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149368.
PR	18-AUG-1999;	99US-0149175.
PR	20-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.

PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 69.3%; Score 1101; DB 21; Length 653;  
Best Local Similarity 67.2%; Pred. No. 3.2e-106;

QY	5	PRV	I	S	D	K	A	S	M	R	S	W	S	R	M	R	A	O	G	K	L	I	G	L	V	P	T	M	G	F	L	H	A	G	H	L	S	L	V	A	Q	A	R	Q	L	S	D	V	V	A	V	S	I	V	N	P	G	64	
Db	346	PEV	I	R	D	K	S	M	R	K	W	S	R	A	M	R	S	O	G	K	T	I	G	L	V	P	T	M	G	Y	L	H	E	G	H	L	S	L	V	R	Q	S	L	A	L	T	D	T	V	V	S	I	V	N	P	G	405		
QY	65	QF	A	P	T	E	D	L	S	T	Y	P	S	D	F	D	G	V	K	K	L	A	S	V	P	G	G	V	D	V	F	H	P	R	N	L	Y	D	G	-----	K	N	G	G	D	V	A	E	A	G	118								
Db	406	QF	S	P	T	E	D	L	S	T	Y	P	S	D	F	S	G	D	L	T	K	A	A	L	S	G	K	V	V	F	N	P	K	N	L	Y	D	G	E	T	K	K	I	N	D	G	G	N	---	G	G	R	462						
QY	119	MV	S	C	V	E	S	G	S	-	G	H	E	S	W	V	R	E	K	L	E	L	G	L	C	G	K	S	R	P	V	F	F	R	G	V	A	T	V	T	K	L	F	N	I	V	E	P	D	V	A	F	G	K	K	D	177		
Db	463	VV	S	C	V	E	E	G	G	L	G	H	E	T	W	I	R	V	E	R	L	E	K	F	C	G	K	S	R	P	V	F	F	R	G	V	A	T	I	V	T	K	L	F	N	I	V	E	P	D	V	A	L	F	G	K	K	D	522
QY	178	QW	R	L	I	Q	R	M	V	R	D	L	F	S	I	K	V	I	G	A	E	I	T	R	D	N	D	G	L	A	M	S	R	N	V	H	L	S	P	E	E	R	E	K	A	L	S	I	N	K	S	L	L	R	A	K	S	237	
Db	523	QW	R	I	I	Q	R	M	V	R	D	L	N	F	G	I	E	V	G	S	D	I	A	R	E	K	D	G	L	A	M	S	R	N	V	R	L	S	D	E	E	R	Q	R	A	L	S	I	S	R	L	A	M	A	K	A	582		
QY	238	A	A	G	D	Q	V	H	C	E	K	L	N	L	V	I	Q	S	V	T	D	A	G	R	I	D	Y	A	E	I	V	D	Q	N	N	L	E	K	V	E	I	Q	I	K	S	P	V	V	F	C	V	A	A	F	G	K	V	R	297
Db	583	S	V	A	E	G	T	N	C	A	E	L	K	D	M	I	I	Q	V	V	G	S	A	G	R	V	D	V	E	I	V	D	Q	E	T	L	E	G	V	E	E	I	K	S	G	V	V	I	C	V	A	A	F	G	T	V	R	642	

QY	298	LIDNMEINLSM 308	
Db	643	LIDNIEINVSL 653	
RESULT 5			
AAG39459			
ID	AAG39459 standard; Protein; 654 AA.		
XX			
AC	AAG39459;		
XX			
DT	18-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 48824.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.		
KW			
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	99US-0139492.
PR	05-MAR-1999;	99US-0123180.	99US-0139454.
PR	09-MAR-1999;	99US-0123548.	99US-0139455.
PR	23-MAR-1999;	99US-0125788.	99US-0139456.
PR	25-MAR-1999;	99US-0126264.	99US-0139457.
PR	29-MAR-1999;	99US-0126785.	99US-0139458.
PR	01-APR-1999;	99US-0127462.	99US-0139459.
PR	06-APR-1999;	99US-0128234.	99US-0139460.
PR	08-APR-1999;	99US-0128714.	99US-0139461.
PR	16-APR-1999;	99US-0129845.	99US-0139462.
PR	19-APR-1999;	99US-0130077.	99US-0139463.
PR	21-APR-1999;	99US-0130449.	99US-0139750.
PR	23-APR-1999;	99US-0130510.	99US-0139763.
PR	23-APR-1999;	99US-0130891.	99US-0139817.
PR	28-APR-1999;	99US-0131449.	99US-0139899.
PR	30-APR-1999;	99US-0132048.	99US-0140353.
PR	30-APR-1999;	99US-0132407.	99US-0140354.
PR	04-MAY-1999;	99US-0132484.	99US-0140695.
PR	05-MAY-1999;	99US-0132485.	99US-0140823.
PR	06-MAY-1999;	99US-0132486.	99US-0140991.
PR	07-MAY-1999;	99US-0132487.	99US-0141287.
PR	11-MAY-1999;	99US-0132863.	99US-0141842.
PR	14-MAY-1999;	99US-0134256.	99US-0142154.
PR	14-MAY-1999;	99US-0134218.	99US-0142055.
PR	14-MAY-1999;	99US-0134219.	99US-0142390.
PR	14-MAY-1999;	99US-0134221.	99US-0142803.
PR	14-MAY-1999;	99US-0134370.	99US-0142920.
PR	18-MAY-1999;	99US-0134768.	99US-0142977.
PR	19-MAY-1999;	99US-0134941.	99US-0143542.
PR	20-MAY-1999;	99US-0135124.	99US-0143624.
PR	21-MAY-1999;	99US-0135353.	99US-0144005.
PR	24-MAY-1999;	99US-0135629.	99US-0144085.
PR	25-MAY-1999;	99US-0136021.	99US-0144086.
PR	27-MAY-1999;	99US-0136392.	99US-0144325.
PR	28-MAY-1999;	99US-0136782.	99US-0144331.
PR	01-JUN-1999;	99US-0137222.	99US-0144332.
PR	03-JUN-1999;	99US-0137528.	99US-0144333.
PR	04-JUN-1999;	99US-0137502.	99US-0144334.
PR	07-JUN-1999;	99US-0137724.	99US-0144335.
PR	08-JUN-1999;	99US-0138094.	99US-0144352.
PR	10-JUN-1999;	99US-0138540.	99US-0144405.
PR	10-JUN-1999;	99US-0138847.	99US-0144086.
PR	14-JUN-1999;	99US-0139119.	99US-0144325.
PR	16-JUN-1999;	99US-0139452.	99US-0144331.
PR	16-JUN-1999;	99US-0139453.	99US-0144332.



PR	13-AUG-1999;	99US-0148684.	
PR	16-AUG-1999;	99US-0149368.	
PR	17-AUG-1999;	99US-0149175.	
PR	18-AUG-1999;	99US-0149426.	
PR	20-AUG-1999;	99US-0149722.	
PR	20-AUG-1999;	99US-0149723.	
PR	20-AUG-1999;	99US-0149929.	
PR	23-AUG-1999;	99US-0149902.	
PR	23-AUG-1999;	99US-0149930.	
PR	25-AUG-1999;	99US-0150566.	
PR	26-AUG-1999;	99US-0150884.	
PR	27-AUG-1999;	99US-0151065.	
PR	27-AUG-1999;	99US-0151066.	
PR	27-AUG-1999;	99US-0151080.	
PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151438.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 69.3%; Score 1101; DB 21; Length 654;			
Best Local Similarity 67.2%; Pred. No. 3.2e-106;			
Matches 209; Conservative 46; Mismatches 46; Indels 10; Gaps 3;			
QY	5	PRVTSDKASMRWSRSMFAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNP	64
Db	347	PEVIRKDSMRKWSRAMRSQGTIGLVPTMGYLGHEGLSLVRQSLALTDVTVSIYVNP	406
QY	65	QFAPTEDLSTYPSDFDGVKKLASVPGGVDVVFHPRNLYDYG-----KNGGSDVAEAGG	118

Db	407	QFSPTEDLSTYPSDFSGDLTKLAALSGGKVVVFNPKNLYDYGGETKKINDGGN--GGR	463
Qy	119	MVSCVESGS-GHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDY	177
Db	464	VVSCVEEGLGHETWIRVERLEKFGCGKSRPVFFRGVATIVTKLFNIVEPDVALFGKKDY	523
Qy	178	QQWELIQRMVRDLDFSIKVIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLLRAKS	237
Db	524	QQWRIIQRMVRDLNFGIEIVGSDIAREKDGLAMSSRNVRLSDEERQALSISRSLAMAKA	583
Qy	238	AAGDQGVHCEKLTNLVIOQVTDAGGRIDYAEIVDQNNLEKVEQIKSPVFCVAAWFGKVR	297
Db	584	SVAEGKTNCAELKDMIIQQVVGSAGRVDYVEIVDQETLEGVEEIKSGVVICVAAWFGTVR	643
Qy	298	LIDNMEINLSM 308	
Db	644	LIDNIEINVSL 654	
RESULT 6			
AAG28281			
ID	AAG28281 standard; Protein; 310 AA.		
XX			
AC	AAG28281;		
XX			
DT	17-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 33440.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.







PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 67.4%; Score 1071; DB 21; Length 299;  
Best Local Similarity 67.2%; Pred. No. 1.4e-103;  
Matches 203; Conservative 46; Mismatches 43; Indels 10; Gaps 3;

QY 14 MRSWSRSMRAQCKLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSVIYVNPQGQFAPTEDLS 73  
Db 1 MRKWSRAMRSQCKTIGLVPTMGYLGHEGHLSLVRQSLALTDVTVVSIVVNPQFSPTEDLS 60

QY 74 TYPSEDFGDGVKKLASVPGGVDVVFHPRNLYDYG-----KNGGDVAEAGGMVSCVESGS 127  
Db 61 TYPSEDFSGDLTKLAALSGGKVVVFNPKNLYDYGGETKKINDGGN---GGRVWSCVEEG 117

QY 128 -GHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLLIQRM 186  
Db 118 LGHETWIRVERLEKXLCGKSRPVFFRGVATIVTKLFNIVEPDVALFGKKDYQQWRLLIQRM 177

QY 187 VRDLDFSIVKIVGAETRDNDGLAMSSRNVHLSPEEREKALSINKSLLRAKSAAGDQGVHC 246  
Db 178 VRDLNFGIEIVGSDIAREKDGLAMSSRNVRLSDEERQALSIISRSLAMAKASVAEGKTNC 237

QY 247 EKLTLNLVIQSVTDAGGRIDYAEIIVDQNNLEKVEQIKSPVVFCAAWFGKVRLIDNMEINL 306  
Db 238 AELKDMIIQQVVGSAGRVDYVEIIVDQETLEGVEEIKSGVVICVAAWFGTVRLIDNIEINV 297

QY 307 SM 308  
Db 298 SL 299

RESULT 9  
AAAY42403  
ID AAY42403 standard; Protein; 333 AA.  
XX

AC AAAY42403;

XX  
DT 02-DEC-1999 (first entry)

XX Amino acid sequence of Rice Pantothenate Synthetase (PS).

XX pantothenate synthetase; cloning; herbicide; biosynthesis;

KW Coenzyme A precursor; pantoate-beta-alanine ligase.

XX  
OS Oryza sativa.

XX

PN WO9942565-A1.  
XX  
PD 26-AUG-1999.  
XX  
PF 02-JUN-1998; 98WO-EP03261.  
XX  
PR 31-MAY-1997; 97GB-0011163.  
PR 27-JUN-1997; 97GB-0013477.  
XX  
PA (AGRE ) HOECHST-SCHERING AGREVO GMEH.  
XX  
PI Abell C, Smith AG, Genschel U, Laber B;  
XX  
DR WPI; 1999-527466/44.  
DR N-PSDB; AAZ20927.

XX New isolated DNA molecule, useful in herbicidal compositions -  
PS Claim 6; Fig 2.2; 64pp; English.  
XX

CC This is the amino acid sequence of the Oryza sativa (rice) Pantothenate  
CC synthetase (PS) enzyme, which is essential in the biosynthesis of  
CC Coenzyme A.

CC The invention provides methods to assay proteins for PS activity, and  
CC the ability of compounds to inhibit PS.

CC The assay compounds which are found to inhibit PS can be used as  
CC herbicides.

CC The nucleotide probe can be used to amplify PS coding sequences from a  
CC chosen organism via the process of polymerase chain reaction.  
CC A number of assays have previously been used to measure PS activity  
CC but are unsuitable for large screen biochemical screening of compounds  
CC to find PS inhibitors. In this invention however the isolation of the  
CC DNA molecule encoding PS enables a host cell to produce the protein  
CC which can then be tested against compounds potentially able to inhibit  
CC the enzyme.

XX  
SQ Sequence 333 AA;

Query Match 58.4%; Score 927.5; DB 20; Length 333;  
Best Local Similarity 59.9%; Pred. No. 1.9e-88;  
Matches 185; Conservative 51; Mismatches 60; Indels 13; Gaps 6;

QY 5 PRVTSDKASMRSMRAQCKLIGLVPTMGFLHAGHLSLV--AQARQLSDVVA--VSTY 60  
Db 27 PEVIRDKAAMRAWSRRRRRAEGKTAVVPTMGYLGHQHLSLISAAAAASADPVAIVVTIY 86

QY 61 VNPQGFAPTEDLSTYPSDFDGVKKLASVPGGVDVVFHPRNLY--DYGKNGGGDVAEAGG 118  
Db 87 VNPQGFAPSEDLATYPSDFDAGDLRKLAST-GVVDAVENPPDLYVRGAGRRGAG---SGG 141

QY 119 MVSCVE--SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKD 176  
Db 142 AISCLEEAAGDHETWVRVERLEKGLCGASRPVFFRGVATIVSKLFNIIIEPDVPVFGKKD 201

QY 177 YQWRLLIQRMVRDLDFSIVKIVGAETRDNDGLAMSSRNVHLSPEEREKALSINKSLLRAK 236  
Db 202 YQWRVILPYWSGLDFGIEIMGSRNCARTDGLAMNSRNVHLSREEGKKALSISRSLVDAR 261

QY 237 SAAGDQGVHCEKLTNLVIQSVTDAGGRIDYAEIIVDQNNLEKVEQIKSPVVFCAAWFGKV 296  
Db 262 TGALKGNTDSKQIKNKIVQTLTETGGQVDYVEIVEQESLVPVEQIDGPVVICVAAWFGKV 321

QY 297 RLIDNMEIN 305  
Db 322 RLIDNIEID 330

RESULT 10  
AAU01245  
ID AAU01245 standard; Protein; 286 AA.  
XX  
AC AAU01245;  
XX



DT 18-JUL-2001 (first entry)  
XX B. subtilis pantothenate synthetase.  
DE Pantothenate synthetase; panC; pantothenate biosynthesis;  
XX vitamin B5; nutritional supplement; panto-compound; pantoate.  
KW Bacillus subtilis.  
XX WO200121772-A2.  
XX 29-MAR-2001.  
XX 21-SEP-2000; 2000WO-US25993.  
XX 21-SEP-1999; 99US-0400494.  
PR 07-JUN-2000; 2000US-0210072.  
PR 28-JUL-2000; 2000US-0221836.  
PR 24-AUG-2000; 2000US-0227860.  
XX (OMNI-) OMNIGENE BIOPRODUCTS.  
PA Yocum RR, Patterson TA, Hermann T, Pero JG;  
XX WPI; 2001-218644/22.  
DR N-PSDB; AAS00989, AAS02314.  
XX New recombinant microorganism which overexpress a Bacillus subtilis  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate -  
XX Example 1; Page 172-173; 292pp; English.  
PS The sequence is B. subtilis pantothenate synthetase, encoded by the  
XX panC gene, an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX SQ Sequence 286 AA;  
Query Match 31.1%; Score 494; DB 22; Length 286;  
Best Local Similarity 36.6%; Pred. No. 5e-43;  
Matches 111; Conservative 51; Mismatches 115; Indels 26; Gaps 3;  
QY 6 RVISDKASMRWSRSMRAQKLI GLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNPQG 65  
Db 2 RQITDISQLKEAIKQYHSEKSGSIGFVPTMGFLHEGHLTLADKARQENDAVIMSFVNPAQ 61  
QY 66 FAPTEDLSTYPSDFDGDVKKLASVPGGVVDVWFHP--RNLYDYGKNGGGDVAEAGGMVSCV 123  
Db 62 FGPNEDEAYPRDIERDAALAE--AGVDILFTPDADHMYPGKN----- 104  
QY 124 ESGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKDYQQWRLI 183  
Db 105 -----VTIHVERRTDVLGCRSREGHFDGVAIVLTCLFNLVKPTRAYFGLKDAQQVAVV 157  
QY 184 QRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLLRKSAAGDQ 243  
Db 158 DGLISDFEMDIELVPDVTVREEDGLAKSSRNVLTAERKEAPKLYRALQTSaelvQAGE 217  
QY 244 VHCEKLTNLVIOQVTDAGGRIDYAEIVDQNNLEKVEIQKSPVFCVAAWFGKVRLLIDNME 303  
Db 218 RDPEAVIKAAKDIIETTSGTIDYVELYSYPELEPVNEIAGKMLAVAVAFSKARLIDNII 277  
QY 304 INL 306

Db 278 IDI 280  
RESULT 11  
ID ABP39476 standard; Protein; 288 AA.  
XX AC ABP39476;  
XX 24-JUL-2002 (first entry)  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4321.  
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX OS Staphylococcus epidermidis.  
XX US6380370-B1.  
XX 30-APR-2002.  
XX 13-AUG-1998; 98US-0134001.  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Doucette-Stamm LA, Bush D;  
PI WPI; 2002-381255/41.  
DR N-PSDB; ABN92021.  
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX Disclosure; SEQ ID 4321; 267pp; English.  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX SQ Sequence 288 AA;  
Query Match 30.4%; Score 482.5; DB 23; Length 288;  
Best Local Similarity 39.1%; Pred. No. 8.1e-42;  
Matches 120; Conservative 45; Mismatches 105; Indels 37; Gaps 6;  
QY 6 RVISDKASMRWSRSMRAQKLI GLVPTMGFLHAGHLSLVAQARQLSDVAVSIYVNPQG 65  
Db 5 KVITTIMQSIYKQHOREGKTIGFVPTMGALHDGHLTMKQSVSENDLTVISIFVNPLO 64  
QY 66 FAPTEDLSTYPSDFDGD---VKKLASVPGGVVDVWFHPRNLYDYGKNGGGDVAEAGMVSC 122  
Db 65 FGPNEDEDAYPRQLDDDDVAAVKKLQ-----VDYVFPSPVDEMYPEELG----- 107  
QY 123 VESGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKDYQQWRL 182  
Db 108 IHLKVGHLAQV-----LEGAQRPGHFEGVTVVVKLFNIVQPDYAYFGKDAQQLAI 159  
QY 183 IQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHLSPEEREKALSINKSLLRKSAAGDG 242  
Db 160 VEKMKDFNLPVHVIGIDIVREKDGGLAKSSRNLYLTSEERKEAKHLYQSLRLAKNLYEAG 219

QY 243 QVHCEKLTNLVIQSVT-----DAGGRIDYAEIVDQNNLEKVEQIKSPVVFCAAWFGKVR 297  
Db 220 -----ERDSNEIIGQIAAYLNKNISGHIDDLGIYSPNLIQQSKIHGRIFFISLAVKFSKAR 275  
QY 298 LIDNMEI 304  
Db 276 LIDNIII 282

RESULT 12  
AAB10683  
ID AAB10683 standard; Protein; 279 AA.  
XX  
AC AAB10683;  
DT 19-JAN-2001 (first entry)  
XX  
DE C. glutamicum panC protein.  
XX  
KW D-pantothenate biosynthesis; panBC; ilvA; L-valine production; ilvD;  
KW dihydroxyacid dehydratase; ilvD; acetohydroxy acid synthase; ilvBN;  
KW isomeroreductase; ilvC; nutrition; threonine dehydratase; panC.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN DE19907567-A1.  
XX  
PD 24-AUG-2000.  
XX  
PF 22-FEB-1999; 99DE-1007567.  
XX  
PR 22-FEB-1999; 99DE-1007567.  
XX  
PA (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
XX  
XX  
PI Eggeling L, Sahm H;  
XX  
DR WPI; 2000-566215/53.  
XX  
XX  
PT Microbial production of L-valine useful as medicine and in nutrition  
PT comprises growing bacteria that have increased activity of selected  
PT genes, e.g. dihydroxyacid dehydratase -  
XX  
PS Example 3; Page 19-20; 24pp; German.  
XX

CC This invention describes a novel method for the microbial production of  
CC L-valine (I) which uses a microorganism in which either the dihydroxyacid  
CC dehydratase (ilvD) activity and/or gene expression or the acetohydroxy  
CC acid synthase (ilvBN) and isomeroreductase (ilvC) activity and/or gene  
CC expression has been increased. (I) is useful in animal and human  
CC nutrition and as a medicine. Increasing expression of the dihydroxyacid  
CC dehydratase, BN and/or C genes results in increased yields of (I),  
CC particularly when used in conjunction with inactivation of genes involved  
CC in synthesis of D-pantothenate e.g. threonine dehydratase (ilvA). This  
CC sequence represents the Corynebacterium glutamicum panC protein which is  
CC used to illustrate the method of the invention.  
XX  
SQ Sequence 279 AA;

Query Match 28.7%; Score 455.5; DB 21; Length 279;  
Best Local Similarity 41.2%; Pred. No. 5.3e-39;  
Matches 121; Conservative 33; Mismatches 93; Indels 47; Gaps 9;

QY 26 KLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGF---APTEDLSTYPSDFDGD 82  
Db 18 KSVGLVPTMGALHSGHASLVKAARAENDTVVASIFVNPLOQFEALGDCDDYRNPRLQDAD 77

QY 83 VKKLASVPGGVDDVVFHPRNLYDYKNGGGDVAE--AGMVSVCVESGSHESWVRVEKLEL 140  
Db 78 LALLEE--AGVDIVFAP-----DVEEYPGGLPLV-----WARTGSIGT 114

QY 141 GLCGKSRPFRGRGATVVTVKLFNIVEPDVAVFGKKDYQQWRLIQRMVRDLDFSIVIGAE 200

Db 115 KLEGASRPGHFHDGVATVVAKLFNLVRPDRAYFGQKDAQQAVIRRLVADLDIPVEIRPVP 174  
QY 201 ITRDNDGLAMSSRNVHLSPEEREKALSINKSL--LRAKSAAGDQGVHCEKLTNLVIQSVT 258  
Db 175 IIRGADGLAESSRNQRLSADQRAQALVLPQVLSGLQRRKAAGEA-----LDIQGAR 225

QY 259 DA-----GGRIDYAEIVDQNNLEKVE---QIKSPVVFCAAWFGKVRRLIDNMEI 304  
Db 226 DTLASADGVRLDHLEIVDPATLEPLEIDGLLTQPALVVGAI FVGVPVRLIDNIEL 279

RESULT 13  
AAB10032  
ID AAB10032 standard; Protein; 279 AA.  
XX  
AC AAB10032;  
DT 02-NOV-2000 (first entry)  
XX  
DE C. glutamicum panC protein.  
XX  
KW L-aspartate-1-decarboxylase; panD; panB; panC; pantothenate synthetase;  
KW ketopantoatehydroxymethyl transferase; D-pantothenic acid; vitamin.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1006192-A2.  
XX  
PD 07-JUN-2000.  
XX  
PF 13-NOV-1999; 99EP-0122650.  
XX  
PR 01-DEC-1998; 98DE-1055313.  
XX  
PA (DEGS ) DEGUSSA-HUELS AG.  
XX  
PI Dusch N, Kalinowski J, Puehler A;  
XX  
DR WPI; 2000-378266/33.  
DR N-PSDB; AAA40273.  
XX

PT Recombinant DNA molecule useful for preparation of vitamin  
PT D-pantothenic acid, comprises Corynebacterium panD gene encoding  
PT aspartate-1-decarboxylase -  
XX

PS Example 2; Page 19-20; 26pp; German.  
XX

CC This invention describes novel recombinant DNA sequences encoding  
CC aspartate-1-decarboxylase (I), also known as panD,  
CC ketopantoatehydroxymethyltransferase (II), also known as panB and  
CC pantothenate synthase (III) also known as panC present in  
CC microorganisms, especially Corynebacterium and Escherichia. (I) is useful  
CC for producing D-pantothenic acid, a vitamin which has important  
CC applications in medicine, cosmetic formulations and animal and human  
CC nutrition. The use of biotechnological and fermentation procedures  
CC allows the production of the desired stereo isoform D-pantothenic acid.  
CC This sequence encodes the Corynebacterium glutamicum pantothenate  
CC synthetase (panC) protein which is described in the method of the  
CC invention.  
XX  
SQ Sequence 279 AA;

Query Match 28.7%; Score 455.5; DB 21; Length 279;  
Best Local Similarity 41.2%; Pred. No. 5.3e-39;  
Matches 121; Conservative 33; Mismatches 93; Indels 47; Gaps 9;

QY 26 KLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGF---APTEDLSTYPSDFDGD 82  
Db 18 KSVGLVPTMGALHSGHASLVKAARAENDTVVASIFVNPLOQFEALGDCDDYRNPRLQDAD 77

QY 83 VKKLASVPGGVDDVVFHPRNLYDYKNGGGDVAE--AGMVSVCVESGSHESWVRVEKLEL 140  
Db 78 LALLEE--AGVDIVFAP-----DVEEYPGGLPLV-----WARTGSIGT 114

Db 78 LALLEE--AGVDIVFAP-----DVEEMYPGGLPLV-----WARTGSIGT 114

QY 141 GLCGKSRPVFFRGVATVVTKLFNIVEPDAVFGKKDYQQWRLIQRMVRDLDFSIKVIGAE 200

Db 115 KLEGASRPGHFDGVATVVAKLFNLRPDRAYFGQKDAQQAVAVIRRLVADLDIPVEIRPVP 174

QY 201 ITRDNDGLAMSSRNVLHLSPEEREKALSINKSL--LRAKSAAGDQGVHCEKLTNLNVIQSVT 258

Db 175 IIRGADGLAESSRNQRLSADQRAQALVLPQVLSGLQRRKAAGEA-----LDIQGAR 225

QY 259 DA-----GGRIDYAEIVDQNNLEKVE---QIKSPVVFCVAAWFGKVRLIDNMEI 304

Db 226 DTLASADGVRLDHLEIVDPATLEPLEIDGLLTQPALVVGAIFFVGPVRLIDNIEL 279

RESULT 14

AAB10034

ID AAB10034 standard; Protein; 279 AA.

XX

AC AAB10034;

DT 02-NOV-2000 (first entry)

XX

DE C. glutamicum panC protein.

XX

KW D-pantothenic acid; panB; panC; ilvD; pantotheanate synthetase;

KW ketopantoathydroxymethyltransferase; dihydroxyaciddehydratase;

KW panBC operon; vitamin.

XX

OS Corynebacterium glutamicum.

XX

PN EP1006189-A2.

XX

PD 07-JUN-2000.

XX

PF 30-NOV-1999; 99EP-0123738.

XX

PR 01-DEC-1998; 98DE-1055312.

XX

PA (DEGS ) DEGUSSA-HUELS AG.

PA (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.

XX

PI Eggeling L, Thierbach G, Sahm H;

XX

XX WPI; 2000-378263/33.

DR

DR N-PSDB; AAA40282.

XX

PT Recombinant Corynebacterium DNA useful for production of pantothenic

PT acid vitamin, comprises panB, panC or ilvD genes encoding enzymes -

XX

XX Example 1; Page 16-17; 27pp; German.

PS

XX This invention describes novel recombinant Corynebacterium DNA (I),

CC present in microorganisms of the Corynebacterium genus and comprising

CC at least one of the panB (ketopantohydroxymethyltransferase), panC

CC (pantothenicacidsynthetase), especially the panBC operon, and/or ilvD

CC (dihydroxyaciddehydratase) genes. (I) is useful for the preparation of

CC pantothenic acid a vitamin which has applications including cosmetics,

CC medicine and human and animal nutrition. The new preparation method using

CC fermentation techniques produces the required stereo-isoform D form of

CC pantothenic acid. This sequence represents the Corynebacterium glutamicum

CC panC protein which is described in the method of the invention.

XX

SQ Sequence 279 AA;

Query Match 28.7%; Score 455.5; DB 21; Length 279;

Best Local Similarity 41.2%; Pred. No. 5.3e-39;

Matches 121; Conservative 33; Mismatches 93; Indels 47; Gaps 9;

QY 26 KLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSYVNPQGF---APTEDLSTYPSDFDGD 82

Db 18 KSVGLVPTMGALHSGHASLVKAARAENDTVVASIFVNPLQFEALGDCDDYRNPRLDAD 77

QY 83 VKKLASVPGGVDVVFHPENLYDYKNGGGDAE--AGGMVSCVESGSHSWVRVEKLEL 140

Db 78 LALLEE--AGVDIVFAP-----DVEEMYPGGLPLV-----WARTGSIGT 114

QY 141 GLCGKSRPVFFRGVATVVTKLFNIVEPDAVFGKKDYQQWRLIQRMVRDLDFSIKVIGAE 200

Db 115 KLEGASRPGHFDGVATVVAKLFNLRPDRAYFGQKDAQQAVAVIRRLVADLDIPVEIRPVP 174

QY 201 ITRDNDGLAMSSRNVLHLSPEEREKALSINKSL--LEAKSAAGDQGVHCEKLTNLNVIQSVT 258

Db 175 IIRGADGLAESSRNQRLSADQRAQALVLPQVLSGLQRRKAAGEA-----LDIQGAR 225

QY 259 DA-----GGRIDYAEIVDQNNLEKVE---QIKSPVVFCVAAWFGKVRLIDNMEI 304

Db 226 DTLASADGVRLDHLEIVDPATLEPLEIDGLLTQPALVVGAIFFVGPVRLIDNIEL 279

RESULT 15

AAG89878

ID AAG89878 standard; Protein; 279 AA.

XX

AC AAG89878;

XX

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 3632.

XX

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

XX

OS Corynebacterium glutamicum.

XX

PN EP1108790-A2.

XX

PD 20-JUN-2001.

XX

PF 18-DEC-2000; 2000EP-0127688.

XX

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX

PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX

DR WPI; 2001-376931/40.

DR N-PSDB; AAH65097.

XX

PT Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT

XX

PS Claim 17; SEQ ID NO: 3632; 246pp + Sequence Listing; English.

XX

CC The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX

SQ Sequence 279 AA;





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 13:34:50 ; Search time 31 Seconds  
(without alignments)  
1859.837 Million cell updates/sec

Title: US-10-033-269-9  
Perfect score: 1588  
Sequence: 1 MAPAPRVISDKASMRWSRS.....AWFGKVRLLIDNMEINLSMNV 310

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1588	100.0	310	12	US-10-033-269-9 Sequence 9, Appli
2	1286	81.0	308	12	US-10-033-269-15 Sequence 15, Appl
3	1164	73.3	311	12	US-10-033-269-7 Sequence 7, Appli
4	1002	63.1	316	12	US-10-033-269-13 Sequence 13, Appl
5	961.5	60.5	325	12	US-10-033-269-2 Sequence 2, Appli
6	959.5	60.4	323	12	US-10-033-269-4 Sequence 4, Appli
7	927.5	58.4	313	12	US-10-033-269-14 Sequence 14, Appl
8	911.5	57.4	296	12	US-10-033-269-11 Sequence 11, Appl
9	455.5	28.7	279	10	US-09-738-626-3632 Sequence 3632, Ap
10	396	24.9	333	15	US-10-156-761-12221 Sequence 12221, A
11	231	14.5	77	15	US-10-113-948-2 Sequence 2, Appli
12	228	14.4	266	10	US-09-738-626-6455 Sequence 6455, Ap
13	178	11.2	79	15	US-10-113-948-1 Sequence 1, Appli
14	94	5.9	659	15	US-10-128-714-8189 Sequence 8189, Ap
15	93.5	5.9	423	15	US-10-127-032-140 Sequence 140, App

16	91	5.7	339	10	US-09-738-626-4524	Sequence 4524, Ap
17	89	5.6	801	15	US-10-156-761-12150	Sequence 12150, A
18	89	5.6	1387	15	US-10-156-761-13000	Sequence 13000, A
19	85.5	5.4	224	10	US-09-993-170-10	Sequence 10, Appl
20	85.5	5.4	385	12	US-10-043-639A-8	Sequence 8, Appli
21	85.5	5.4	630	12	US-09-882-227-460	Sequence 460, App
22	85.5	5.4	715	15	US-10-156-761-9360	Sequence 9360, Ap
23	85	5.4	479	11	US-09-934-455-210	Sequence 210, App
24	84.5	5.3	223	10	US-09-993-170-3	Sequence 3, Appli
25	84.5	5.3	224	10	US-09-993-170-13	Sequence 13, Appl
26	84.5	5.3	565	14	US-10-062-254-92	Sequence 92, Appl
27	84	5.3	706	15	US-10-270-333-153	Sequence 153, App
28	83.5	5.3	2756	15	US-10-331-061-7	Sequence 7, Appli
29	82.5	5.2	5147	12	US-10-174-677-4	Sequence 4, Appli
30	82	5.2	583	10	US-09-758-269-2	Sequence 2, Appli
31	82	5.2	1090	9	US-09-833-435A-5	Sequence 5, Appli
32	82	5.2	1090	12	US-10-375-720-5	Sequence 5, Appli
33	82	5.2	1091	11	US-09-262-126C-4	Sequence 4, Appli
34	82	5.2	1091	15	US-10-245-803-4	Sequence 4, Appli
35	82	5.2	5183	12	US-10-107-521-1	Sequence 1, Appli
36	81.5	5.1	223	10	US-09-993-170-7	Sequence 7, Appli
37	81.5	5.1	361	12	US-10-166-225A-65	Sequence 65, Appl
38	81.5	5.1	691	12	US-10-321-802-22	Sequence 22, Appl
39	80.5	5.1	403	10	US-09-864-921-176	Sequence 176, App
40	80.5	5.1	900	9	US-09-815-242-13381	Sequence 13381, A
41	80.5	5.1	900	9	US-09-815-242-13663	Sequence 13663, A
42	80.5	5.1	1009	10	US-09-864-921-107	Sequence 107, App
43	80.5	5.1	4349	12	US-10-174-677-76	Sequence 76, Appl
44	80.5	5.1	4349	12	US-09-970-944-4	Sequence 4, Appli
45	80.5	5.1	4349	12	US-09-970-944-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1

US-10-033-269-9  
; Sequence 9, Application US/10033269  
; Publication No. US20030167503A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutteridge, Steve  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Orozco, Buddy  
; TITLE OF INVENTION: Plant Genes Encoding Pantothenate Synthetase  
; FILE REFERENCE: BBI446 US NA  
; CURRENT APPLICATION NUMBER: US/10/033,269  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/247,938  
; PRIOR FILING DATE: 2000-11-13  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 9  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-033-269-9

Query Match		100.0%;	Score 1588;	DB 12;	Length 310;
Best Local Similarity		100.0%;	Pred. NO. 1.4e-158;		
Matches 310;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAPAPRVISDKASMRWSRSMRAQKLI	GLVPTMGFLHAGHLSLVAQARQLSDVAVSIY	60	
Db	1	MAPAPRVISDKASMRWSRSMRAQKLI	GLVPTMGFLHAGHLSLVAQARQLSDVAVSIY	60	
QY	61	VNPGQFAPTEDLSTYPSDFDGDVKLASV	PGGVVDVVFHPRNLVDYGNKGGDVAEAGMV	120	
Db	61	VNPGQFAPTEDLSTYPSDFDGDVKLASV	PGGVVDVVFHPRNLVDYGNKGGDVAEAGMV	120	
QY	121	SCVESGSGHESWVRVEKLELGLCGKSRP	VFRGVATVVTKLFNIVEPVAVFGKDYQQW	180	
Db	121	SCVESGSGHESWVRVEKLELGLCGKSRP	VFRGVATVVTKLFNIVEPVAVFGKDYQQW	180	





Db 64 YNPSQFAPTEDLATYPSDLAGLRKLAST-GAVHAVFNPDLYHRGAASVGRRAEPAG 122  
QY 117 GGMVSCVES-GSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKK 175  
Db 123 AAASSCLEAGGDGHETWIRVERLEKGLCGASRPVFFRGVATVVAKLFNVVEPDVAMFGKK 182  
QY 176 DYQWRLIQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLRA 235  
Db 183 DYQWRLICRMVRDLDFAVEIIGAEIVREADGLAMSSRNVHLSPEEREKALSISRSLNA 242  
QY 236 KSAAGDQGVHCEKLTNLVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVVFCAAWFG 294  
Db 243 RTAALNNSASASEHIKDQIVQTLTEAGGRVDYVEIVQESILVPVETIDRPVVICVAAWFG 302  
QY 295 KVRLLDNMEINL 306  
Db 303 KVRLLDNIEIHI 314

RESULT 5  
US-10-033-269-2  
; Sequence 2, Application US/10033269  
; Publication No. US20030167503A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutteridge, Steve  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Orozco, Buddy  
; TITLE OF INVENTION: Plant Genes Encoding Pantothenate Synthetase  
; FILE REFERENCE: BB1446 US NA  
; CURRENT APPLICATION NUMBER: US/10/033,269  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/247,938  
; PRIOR FILING DATE: 2000-11-13  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 2  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-033-269-2

Query Match 60.5%; Score 961.5; DB 12; Length 325;  
Best Local Similarity 62.0%; Pred. No. 1.5e-92;  
Matches 191; Conservative 49; Mismatches 61; Indels 7; Gaps 4;

QY 5 PRVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSD---VVAVSIYV 61  
Db 11 PEVIRDKAAMRAWSRRRRRAEGKAVALVPTMGFLHEGHLSLSAAVAASAGPIAVVSIYV 70  
QY 62 NPGQFAPTEDLSTYPSDFDGDVKKLASVPGGVDVVFHPRNLYDYGNKGGDVAEA-GGMV 120  
Db 71 NPSQFAPTEDLATYPSDFAGDLRKLAAT-GVAAAVFCPPDLYVRGSADRPSPAAGSGGAV 129  
QY 121 SCVESGSG--HESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQ 178  
Db 130 SCLEDAGGHAHETWIRVERLEKGLCGSSRPVFFRGVATVVAKLFNIVEPDVAVFGKKDYQ 189  
QY 179 QWRLIQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLRAKSA 238  
Db 190 QWRVICRMVRDLDFAIQIVGSEVVREADGLAMSSRNVLSEEDRKKALSISRSLVDARTA 249  
QY 239 AGDQGVHCEKLTNLVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVVFCAAWFGKVR 298  
Db 250 ALSGNSRSQEIKDQIVRTITEAGGQVDYVEIVQESILVPVERMDRPCVICVAAWFGKVR 309  
QY 299 IDNMEINL 306  
Db 310 IDNIEIHV 317

RESULT 6  
US-10-033-269-4

; Sequence 4, Application US/10033269  
; Publication No. US20030167503A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutteridge, Steve  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Orozco, Buddy  
; TITLE OF INVENTION: Plant Genes Encoding Pantothenate Synthetase  
; FILE REFERENCE: BB1446 US NA  
; CURRENT APPLICATION NUMBER: US/10/033,269  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/247,938  
; PRIOR FILING DATE: 2000-11-13  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 4  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-033-269-4

Query Match 60.4%; Score 959.5; DB 12; Length 323;  
Best Local Similarity 62.0%; Pred. No. 2.4e-92;  
Matches 191; Conservative 48; Mismatches 62; Indels 7; Gaps 4;  
QY 5 PRVISDKASMRWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSD---VVAVSIYV 61  
Db 9 PEVIRDKAAMRAWSRRRRRAEGKAVALVPTMGFLHEGHLSLVSAAVASAGPVAVVSIYV 68  
QY 62 NPGQFAPTEDLSTYPSDFDGDVKKLASVPGGVDVVFHPRNLYDYGNKGGDVAEA-GGMV 120  
Db 69 NPSQFAPTEDLATYPSDFAGDLGKLAAT-GVAAAVFCPPDLYVRGSADRPSPAAGSGAV 127  
QY 121 SCVESGSG--HESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQ 178  
Db 128 SCLEDAGGHAHETWIRVERLEKGLCGSSRPVFFRGVATVVAKLFNIVEPDVAVFGKKDYQ 187  
QY 179 QWRLIQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVHLSPEEREKALSINKSLRAKSA 238  
Db 188 QWRVICRMVRDLDFAIQIVGSEVVREADGLAMSSRNVLSEEDRKKALSISRSLVDARTA 247  
QY 239 AGDQGVHCEKLTNLVIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSPVVFCAAWFGKVR 298  
Db 248 TLGSGNSRSQEIKDQIVRTITEAGGQVDYVEIVQESILVPVERMDRPCVICVAAWFGKVR 307  
QY 299 IDNMEINL 306  
Db 308 IDNIEIHV 315

RESULT 7  
US-10-033-269-14  
; Sequence 14, Application US/10033269  
; Publication No. US20030167503A1  
; GENERAL INFORMATION:  
; APPLICANT: Gutteridge, Steve  
; APPLICANT: Harvell, Leslie T.  
; APPLICANT: Orozco, Buddy  
; TITLE OF INVENTION: Plant Genes Encoding Pantothenate Synthetase  
; FILE REFERENCE: BB1446 US NA  
; CURRENT APPLICATION NUMBER: US/10/033,269  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/247,938  
; PRIOR FILING DATE: 2000-11-13  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 14  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-033-269-14

Query Match 58.4%; Score 927.5; DB 12; Length 313;  
Best Local Similarity 59.9%; Pred. No. 5.4e-89;







```
US-10-113-948-1
; Sequence 1, Application US/10113948
; Publication No. US20030082773A1
; GENERAL INFORMATION:
; APPLICANT: Blundell, Tom L
; APPLICANT: Abell, Christopher
; APPLICANT: von Delft, Frank
; TITLE OF INVENTION: Crystal Structure
; FILE REFERENCE: 620-194
; CURRENT APPLICATION NUMBER: US/10/113,948
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US 09/659,759
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: PCT/GB01/04067
; PRIOR FILING DATE: 2001-09-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-113-948-1

Query Match      11.2%; Score 178; DB 15; Length 79;
Best Local Similarity 51.3%; Pred. No. 5.8e-11;
Matches 39; Conservative 10; Mismatches 25; Indels 2; Gaps 1;

QY 24 QGKLIGLVPTMGFLHAGHLSLVAQARQLSPVVAVSIYVNPQGFAPTEDLSTYPSDFDGDV 83
Db 1 EGKRVALVPTMGNLHDGHMKLVDEAKARADVVVVVSIFVNPQMQFDRPEDLARYPRTLQEDC 60

QY 84 KKLASVPGGVVDVVFHP 99
Db 61 EKLNK--RKVDLVFAP 74

RESULT 14
US-10-128-714-8189
; Sequence 8189, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8189
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8189

Query Match      5.9%; Score 94; DB 15; Length 659;
Best Local Similarity 18.9%; Pred. No. 1;
Matches 66; Conservative 52; Mismatches 107; Indels 124; Gaps 14;
```

```
QY 10 DKASMRSWSRMRAQKGKLIGLVPTMGFLH-----AGHLSLVAQARQLSDVVAVSIY 60
Db 94 EQVSVRGWKQSKAASNRDGGLESLIAFLEKKLTASDSKAGARARITKSRVEGDSVIVSI- 152
QY 61 VNPQGFAPTEDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNLYDYGNKGGG-----D 112
Db 153 -RPGLLDRMLQMN-----GFSFAGAPLTIEAYGPSANGMTDQPMLSD 193
QY 113 VAEAGGMVSCVESGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVF 172
Db 194 IAQNGGAPSTADTKS-----KM-----TAIL 214
QY 173 GKDYQQWRLIQRMVRDLDFSIKVIGAEITRNDGLAMSSRNVHLSPEEREKAL----- 226
Db 215 SKRYIQQTKL-----LDLS-----KLGSDPDLLAMGIFNSTSTESKFFPALMKVWEM 261
QY 227 SINKSLLRAKS-----AAGDGOVHCEKLTNLVIOQSVTDAGGRIDYAEIVDQN----- 273
Db 262 SFDNSTTRREAVESVSLADNQLANIAVVTSLAQITPD-----LRNLDLSNNNFKDAQSLIG 317
QY 274 -----NLEKVEQIKSPV-----VF--CVAAWFGKVRLIDNMEINLSMNV 310
Db 318 WRWKFRNLEFLDLTGTPFSADPTFKETMMKWYPKLRVLNNVEVRTAEEI 366

RESULT 15
US-10-127-032-140
; Sequence 140, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Banger, M. Gita
; APPLICANT: Lory, Stephen
; APPLICANT: Greenberg, Everett Peter
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
; TITLE OF INVENTION: BIOFILM FORMATION
; FILE REFERENCE: UIZ-070CP
; CURRENT APPLICATION NUMBER: US/10/127,032
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,190
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/344,142
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 140
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-127-032-140

Query Match      5.9%; Score 93.5; DB 15; Length 423;
Best Local Similarity 22.5%; Pred. No. 0.59;
Matches 54; Conservative 31; Mismatches 98; Indels 57; Gaps 9;
```

```
QY 70 EDLSTYPSDFDGDVKKLASVPGGVVDVVFHPRNL-YDYGKN-----GGDVAAEAGGMVSCVE 124
Db 50 EQISIPGRDIDEPV--LHHGRGRQTVVHPGNKEFTAGEHIAIAPSGGGGGGGGKAS--N 105
QY 125 SGSGHESWVRVEKLELGLCGKSRPVFFRGVATVVTKLFNIVEPDVAVFGKKDYQQWRLIQ 184
Db 106 SGEGMDDFV-----FQITQEEFLDFMFEDLELPNLVK 137
QY 185 RMVRDLDFSIVIGAEITRNDGLAMS-----SRNVHLSPEEREKALSINKSLRA 235
Db 138 RHITGTD-TFKTVRAGISNDGNPSRINIVRTLRSAHARRIALSGGSRAKLRALKELERI 196
QY 236 KSAAGDGOVHCEKLTNLVIOQSVTDAGGRIDYAEIVDQNNLEKVEQIKSP-----VVFCV 289
Db 197 KREPDNLGDIQELE----LEIAKLRARIDRVPFLDTFDLKYNNLLVKQPNPTSKAVMFCL 252
```

Search completed: December 15, 2003, 13:40:58  
Job time : 32 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 13:35:45 ; Search time 21 Seconds  
(without alignments)  
624.589 Million cell updates/sec

Title: US-10-033-269-9  
Perfect score: 1588  
Sequence: 1 MAPAPRVISDKASMRWSRS.....AWFGKVRLLDNMEINLSMNV 310

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	567	35.7	294	4	US-09-107-532A-5153
2	514.5	32.4	283	4	US-09-328-352-7916
3	482.5	30.4	288	4	US-09-134-001C-4321
4	481	30.3	473	4	US-09-252-991A-26805
5	455.5	28.7	279	3	US-09-318-794A-3
6	455.5	28.7	279	3	US-09-318-793A-5
7	103.5	6.5	421	2	US-08-576-626A-53
8	103.5	6.5	1114	2	US-08-576-626A-31
9	100.5	6.3	379	4	US-09-107-532A-4933
10	93.5	5.9	426	4	US-09-252-991A-17427
11	93	5.9	422	4	US-09-679-279-17
12	90.5	5.7	423	4	US-09-328-352-4686
13	90.5	5.7	657	3	US-07-705-490-14
14	90.5	5.7	657	3	US-07-751-891B-14
15	89.5	5.6	373	4	US-09-328-352-7186
16	89	5.6	565	4	US-09-252-991A-16764
17	87.5	5.5	760	4	US-09-252-991A-27790
18	87	5.5	566	4	US-09-252-991A-25622
19	86.5	5.4	1250	3	US-08-938-291A-9
20	86.5	5.4	1250	3	US-09-589-619-9
21	84.5	5.3	223	4	US-09-336-946B-6
22	84.5	5.3	675	4	US-09-252-991A-29071
23	83.5	5.3	2756	1	US-08-375-709-11
24	83.5	5.3	2756	1	US-08-752-929-11
25	83.5	5.3	2756	3	US-09-090-793-7
26	83.5	5.3	2756	4	US-09-231-899-7
27	82.5	5.2	500	4	US-09-328-352-4471

28	82	5.2	497	3	US-08-724-466B-4	Sequence 4, Appli
29	82	5.2	497	4	US-08-882-164D-4	Sequence 4, Appli
30	82	5.2	1090	3	US-09-346-237-5	Sequence 5, Appli
31	81.5	5.1	15281	2	US-08-471-119A-2	Sequence 2, Appli
32	81	5.1	398	4	US-09-242-859A-4	Sequence 4, Appli
33	81	5.1	398	4	US-09-242-859A-8	Sequence 8, Appli
34	81	5.1	436	4	US-09-679-279-4	Sequence 4, Appli
35	80.5	5.1	236	1	US-08-684-862-5	Sequence 5, Appli
36	80.5	5.1	930	3	US-09-283-763-2	Sequence 2, Appli
37	80.5	5.1	930	4	US-09-574-912-2	Sequence 2, Appli
38	80.5	5.1	1287	4	US-09-252-991A-29606	Sequence 29606, A
39	79.5	5.0	312	4	US-09-252-991A-17305	Sequence 17305, A
40	79	5.0	187	3	US-09-095-855-203	Sequence 203, App
41	79	5.0	187	4	US-09-205-426-203	Sequence 2, Appli
42	79	5.0	506	2	US-08-929-501-2	Sequence 2, Appli
43	79	5.0	506	3	US-09-140-177-2	Sequence 2, Appli
44	79	5.0	506	3	US-09-397-979-2	Sequence 2, Appli
45	79	5.0	590	2	US-08-929-501-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1  
US-09-107-532A-5153  
; Sequence 5153, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5153:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 294 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...294  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5153:

US-09-107-532A-5153







```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.O1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998194e
US-08-576-626A-53

Query Match      6.5%; Score 103.5; DB 2; Length 421;
Best Local Similarity 21.6%; Pred. No. 0.0064;
Matches 74; Conservative 39; Mismatches 109; Indels 121; Gaps 17;

QY 14 MRSWSRSMRAQKLIQLVP-TMGFLHAGHLSLVAQARQLSDVAVSIYVNPQGQFAPTEDL 72
Db 1 MRVVFSSMASKSHLFGVLPLAWAFRAAGH-----EVRVVASP---ALTEDI 43

QY 73 STYPSDFDGDVKKLASVPGGVVVFHPRNLVDYKNGGGDVAEAGMVSCVSGSGHESW 132
Db 44 T-----AAGLTAVPVGTDV-----DLVDFMTHAGHDIIDYVRSDFSERDPATLTW 89

QY 133 VRVEKLE-----LGLCGKSRP-----VFFRGVATVVTKLFN- 163
Db 90 EHLRGMQTVLTPTFYALMSPDTLIEGMVSFCRKWRPDLVIWEPLTFAAPIAGAVTGTGPHA 149

QY 164 --IVEPDVAVFGKKDY-----LGLCGKSRP-----VFFRGVATVVTKLFN- 182
Db 150 RLLWGPDIITTRARQNFLGLLDPQPEEHREGPLAEWLTTWLEKYGGPAFDEEVVVGQWTID 209

QY 183 -IQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHSPPEEREKALSINKSLRAKSAAGD 241
Db 210 PAPAAIR-LDTGLKTVGMRYV-DYNGPSVVPWLHDEPERRRRVCLTLGIS-SRENSI--- 263

QY 242 GQVHCEKLTNLVIQSVTDAGGRIDYAEIV---DQNNLEKVEQI 281
Db 264 GQVSIIEELLGAV-----GDVD-AEIIATFDAQQLEGVANI 297

RESULT 8
US-08-576-626A-31
; Sequence 31, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.O1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998194e
US-08-576-626A-31

Query Match      6.5%; Score 103.5; DB 2; Length 1114;
Best Local Similarity 21.6%; Pred. No. 0.03;
Matches 74; Conservative 39; Mismatches 109; Indels 121; Gaps 17;

QY 14 MRSWSRSMRAQKLIQLVP-TMGFLHAGHLSLVAQARQLSDVAVSIYVNPQGQFAPTEDL 72
Db 361 MRVVFSSMASKSHLFGVLPLAWAFRAAGH-----EVRVVASP---ALTEDI 403

QY 73 STYPSDFDGDVKKLASVPGGVVVFHPRNLVDYKNGGGDVAEAGMVSCVSGSGHESW 132
Db 404 T-----AAGLTAVPVGTDV-----DLVDFMTHAGHDIIDYVRSDFSERDPATLTW 449

QY 133 VRVEKLE-----LGLCGKSRP-----VFFRGVATVVTKLFN- 163
Db 450 EHLRGMQTVLTPTFYALMSPDTLIEGMVSFCRKWRPDLVIWEPLTFAAPIAGAVTGTGPHA 509

QY 164 --IVEPDVAVFGKKDY-----LGLCGKSRP-----VFFRGVATVVTKLFN- 182
Db 510 RLLWGPDIITTRARQNFLGLLDPQPEEHREGPLAEWLTTWLEKYGGPAFDEEVVVGQWTID 569

QY 183 -IQRMVRDLDFSIVKIGAEITRDNDGLAMSSRNVLHSPPEEREKALSINKSLRAKSAAGD 241
Db 570 PAPAAIR-LDTGLKTVGMRYV-DYNGPSVVPWLHDEPERRRRVCLTLGIS-SRENSI--- 623

QY 242 GQVHCEKLTNLVIQSVTDAGGRIDYAEIV---DQNNLEKVEQI 281
Db 624 GQVSIIEELLGAV-----GDVD-AEIIATFDAQQLEGVANI 657

RESULT 9
US-09-107-532A-4933
; Sequence 4933, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII

```



```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4933:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...379
; SEQUENCE DESCRIPTION: SEQ ID NO: 4933:
US-09-107-532A-4933

Query Match
Best Local Similarity 23.3%; Score 100.5; DB 4; Length 379;
Matches 74; Conservative 34; Mismatches 97; Indels 113; Gaps 17;

QY 50 QLSDV-VAVSIYVNPQGFAPTEDLSTYPSDFDGDVKKLASVPGGVDV--VFHP----- 99
Db 61 EMSDLAIKTGAYLS--QLLKVEDAQIVSSASAGIAQSVAAALIGQDMYHVYHPYTEKIAR 118
QY 100 RNL-----YDYG-----KNGGGDVAEAGGMVSCVESGSGHESWVRVEKLELGLCGK 145
Db 119 REIIPKGNVDYGTPEVWVMAQGGKVIEAGYANMCTPE--HVSMMITEKTAALLYIK 175
QY 146 SRPVFFRGVATV-----VTK-----LFNIVEPDV----- 169
Db 176 SHHAVQKSLTVAEMVEVAKAHLRLPLIVDAAAEEDLFKYSEMVGVDLVIYSGAKAIEGPSA 235
QY 170 -AVFGKKDYQQWRLIQRMVRDLDFSIVIG--ABITRDNDGLAMSSRNVHLSPEEREKAL 226
Db 236 GLVIGKKYIQWIRLOS-----KGIGRAMKIGKN-----IL 267
QY 227 SINKSL---LRAKSAAGDQGVH-----CEKLTNLIQSVTDAGGRIDYAEIV-----D 271
Db 268 GFTQALEDYLKNGSESGDSMKQRLAPFVNAINQLPDLSAKIVQDGAGRDIYRASVTINGD 327
QY 272 QNNLEKVEQIK--SPVVF 287
Db 328 KSAKEVIQELKARNPAVY 345

RESULT 10
US-09-252-991A-17427
; Sequence 17427, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
```

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17427
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17427

Query Match
Best Local Similarity 5.9%; Score 93.5; DB 4; Length 426;
Matches 54; Conservative 31; Mismatches 98; Indels 57; Gaps 9;

QY 70 EDLSTYPSDFDGDVKKLASVPGGVDVVFHPRNL-YDYGKN-----GGDVAEAGGMVSCVE 124
Db 53 EQISIPGRDIDEPV--LHHGRGGRQTVVHPGNKEFTAGEHIARPSGGGGRGGKAS--N 108
QY 125 SSGHESWVRVEKLELGLCGKSRPVFFRGVATVTKLFNIVEPDVAVFGKKDYQQWRLIQ 184
Db 109 SGEGMDDFV-----FOITQEEFLDFMEDLELPNLVK 140
QY 185 RMVRDLDFSIVKIVGAETIRNDGLAMS-----SRNVHLSPEEREKALSINKSLLRA 235
Db 141 RHITGTD-TFKTVRAGISNDGNPSRINIVRTLSAHARRIALSGGSRAKLRAALKELERI 199
QY 236 KSAAGDQGVHCEKLTNLIQSVTDAGGRIDYAEIVDQNNLEKVEQIKSP-----VVFV 289
Db 200 KREPDNLGDIQELE----LEIAKLRARIDRVPFLLDTFDLKYNLLVQPNPTSKAVMFCL 255

RESULT 11
US-09-679-279-17
; Sequence 17, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Volchegursky, Yanina
; TITLE OF INVENTION: Recombinant Megalomycin Biosynthetic
; TITLE OF INVENTION: Genes and Uses Thereof
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Micromonospora megalomicea
US-09-679-279-17

Query Match
Best Local Similarity 5.9%; Score 93; DB 4; Length 422;
Matches 64; Conservative 32; Mismatches 99; Indels 70; Gaps 14;

QY 14 MRSWSRSMRAQKKLIGLVP-TMGFLHACHLSLVAQARQLSDVVAVSVIYVNPQGFAPTE 72
Db 1 MRVVFSSMASKSHLFGVLPLAWAFRAAGH-----EVRVVASP---ALTTDDI 43
QY 73 STYPSDFDGDVKKLASVPGGVDVVFHPRNLVDYGNKGGDVAEAGGMVSCVESGSGHESW 132
Db 44 T-----AAGLTAVPVGTDV-----DLVDFMTHAGYDIDYVRSLDFSERDPATSTW 89
QY 133 VRVEKLELGLCGKSRPVFFRGVATVVTKLKFNIVEPDVAVFGKKDY-QQWRLIQRMVRDL 190
Db 90 DHL-----LGMQTVLTPTFYALMSPDSLVEGMISFCRSWR-----PDW 127
QY 191 DFSIVKIVGAETIRNDGLAMSSR--NVHLSPEEREKALSINKSLLRAKSAAGDQGVHCE- 247
Db 128 SSGPQTFAAASIAATVTGVAHARLLWGPDITVRARQKFL----GLLPQOPAA-----HRED 178
```

QY 248 KLTNLVIOQVTDAGGRI--DYAEIV 270  
Db 179 PLAELWTSVERFGGRVPQDVEELV 203

RESULT 12  
US-09-328-352-4686  
; Sequence 4686, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4686  
; LENGTH: 423  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4686

Query Match 5.7%; Score 90.5; DB 4; Length 423;  
Best Local Similarity 22.3%; Pred. No. 0.19;  
Matches 69; Conservative 43; Mismatches 82; Indels 115; Gaps 20;

QY 8 ISDKASMRWSRSMRAQGLIGLVPTMGFLHAGHLSLVAQARQLSDVVAVSIY-----V 61  
Db 145 LSPNAS-GGYAEQVLAQSSLIFEVPN-----GLDADSAAMTEPMAVALHAVRRSRV 194  
QY 62 NPGQFAPTEDLSTYP-----SDFDGDVKKLASVPGGVDDVVFHPRNL 102  
Db 195 KTGE--PAIVIGCGPVGLGVILMLKAAGVKTVVASDFSPNRRKLAE-QCGADIVVDPKET 251  
QY 103 YDYG--KNGG--GDVAEAGGMVSCVESGSG-----HESWVRVEKLELGLCGKS 146  
Db 252 SPFANWKEFGLGNVSDA-----INMGMLGFLDKLQATRLPWWHGWRMTDK--LGALPK- 302  
QY 147 RPVFFRGVATVVTKLENIPEPDVAVFGKKDYQQWRLIQRMVRDLDFSIVIGAEITRDND 206  
Db 303 RPVIFECGV-----PGV-----LQIIIEGAPLFSRIVGV----- 332  
QY 207 GLAMSSRNVHLSPEEREKALSINKSL-----LRAKSA---AGDGOVHCEKL-TNL 252  
Db 333 GVCMQSDKI-----EPALAINKELEIQVLGYTPLEFRDALHMAIEGKVNCSPLITGV 385  
QY 253 V-IQSVTDA 260  
Db 386 VGLEGVTNA 394

RESULT 13  
US-07-705-490-14  
; Sequence 14, Application US/07705490  
; Patent No. 6107025  
; GENERAL INFORMATION:  
; APPLICANT: Caskey, C. T.  
; APPLICANT: Nelson, David L.  
; APPLICANT: Pieretti, Maura  
; APPLICANT: Warren, Stephen T.  
; APPLICANT: Oostra, Ben A.  
; TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thomas D. Paul  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: U.S.A.  
; ZIP: 77010-3095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/705,490  
; FILING DATE: 19910708  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul, Thomas D.  
; REGISTRATION NUMBER: 32,714  
; REFERENCE/DOCKET NUMBER: D-5350  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713/651-5325  
; TELEFAX: 713/651-5246  
; TELEX: 762829  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 657 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: YES  
; FRAGMENT TYPE: C-terminal  
US-07-705-490-14

Query Match 5.7%; Score 90.5; DB 3; Length 657;  
Best Local Similarity 19.5%; Pred. No. 0.38;  
Matches 61; Conservative 34; Mismatches 91; Indels 127; Gaps 11;

QY 34 MGFLHAGHLSLVAQARQLSDVVAVSIYVNPQGAFAPTEDLST---YPSDFDGDVKKLASVP 90  
Db 295 MGLAIGHGANIQQARKVPGVTAIDL-----DEDTCTFHIYGEDQDAVKKARSFLE 345  
QY 91 GGVDVVFHPRNLYDYGNKG--GDVAEAGGMVSC----- 122  
Db 346 FAEDVIQVPRNLVWIGKNGKLIQEIVDKSGVVRVRIEAENKKNVPQEEEEIMPPNLSNN 405  
QY 123 -----VESGSGHESVVRVEKLELGLCGKSRPVFRGV-----ATVVTKL 161  
Db 406 SRVGNPAPEKKHLDIKENSTHFSQPNSTKVQGMV---PFVFGTKDSIANATVLL-- 459  
QY 162 FNIVEPDVAVFGKKDYQQWRLIQRMVRDLDFSIVIGAE-----ITRDNDG 207  
Db 460 -----DYHLNLYLKEVDQLRL-ERL--QIDEQLRQIGASSRPPPNRTDKEKSYVTDDGQG 510  
QY 208 LAMSSR----- 570  
Db 511 MGRGSRPYRNRHGRRGPGYTGTSNSEASNASETESDHRDELSDWSLAPTEERESEFLRR 570  
QY 231 SLLRAKSAAGDQG 243  
Db 571 GDGRRRGGGGRGQ 583

RESULT 14  
US-07-751-891B-14  
; Sequence 14, Application US/07751891B  
; Patent No. 6180337  
; GENERAL INFORMATION:  
; APPLICANT: Caskey, C. T.  
; APPLICANT: Nelson, David L.  
; APPLICANT: Pieretti, Maura  
; APPLICANT: Warren, Stephen T.  
; APPLICANT: Oostra, Ben A.  
; APPLICANT: Fu, Ying-hui  
; TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thomas D. Paul  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: U.S.A.

; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/751,891B
; FILING DATE: 29-Aug-1991
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; FRAGMENT TYPE: C-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-07-751-891B-14

Query Match 5.7%; Score 90.5; DB 3; Length 657;
Best Local Similarity 19.5%; Pred. No. 0.38;
Matches 61; Conservative 34; Mismatches 91; Indels 127; Gaps 11;

QY 34 MGFLHAGHLSLVAQARQLSDVWAVSIYVNPQGFAPTEDLST---YPSDFDGDVKKLASVP 90
Db 295 MGLAIGTHGANIQARKVPGVTAIDL-----DEDTCTFHIYGEDQDAVKKARSFLE 345

QY 91 GGVDVVFHPRNLYDYGNKG--GDVAEAGGMVSC----- 122
Db 346 FAEDVIQVPRNLVVGKNGKLIQEIYVDKSGVVRVRIEAEENEKNVPQEEIMPNSLPSNN 405

QY 123 -----VESGSGHESWVRVEKLELGLCGKSRPVFRGV-----ATVVTKL 161
Db 406 SRVGNPAPEKXHLDIKENSTHFSQPNSTKVQRGMV---PFVFGTKDSIANATVLL-- 459

QY 162 FNIVEPDVAVFGKKDYQQWRLIQRMVRDLDFSIKIVIGAE-----ITRDNDG 207
Db 460 -----DYHLNLYKEVDQLRL-ERL--QIDQLRQIGASSRPPPNRTDKESYVTDGQG 510

QY 208 LAMSSR-----NVHLSPEEREKALSINK 230
Db 511 MGRGSRPYRNRGHRGRRGPGYTSGTNSSEASNASETESDHRDELSWSLAPTEERESFLRR 570

QY 231 SLLRAKSAAGDQ 243
Db 571 GDGRRRGGGGRGQ 583

RESULT 15
US-09-328-352-7186
; Sequence 7186, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7186
; LENGTH: 373
; TYPE: PRT

; ORGANISM: Acinetobacter baumannii
; US-09-328-352-7186

Query Match 5.6%; Score 89.5; DB 4; Length 373;
Best Local Similarity 19.9%; Pred. No. 0.2;
Matches 64; Conservative 48; Mismatches 106; Indels 103; Gaps 14;

QY 2 APAPRVISDKA--SMRSWSRSMRAQKGLIGLVPTMGFLHAGHLSLVAQARQLSDVAVSI 59
Db 25 AALPTAVLEKAQQELLDW-----QKG-----GLSINEMSHRSADYVAVA- 63

QY 60 YVNPQGQFAPTEDLSTYPSDFDGDVKKLASVP-----GGVDVVFH--PRNLYDYGKN 108
Db 64 -----EKAEDLRKLMNIPENYKVLFLQGGASLQFSAIPLNL--LGKN 104

QY 109 GGDVABEAG-----GMVSCVESGSGHESWVRV-EKLELGLCGKSRPVFRGV 154
Db 105 NKADYIHTGIWSEKALKAEAKRYGDIVNVEAGIKVDGKFAISEQSEWNLSDDAAYVHYADN 164

QY 155 ATVVTKLENIVEPDVAVFGKKDYQQWRLIQRMVRDLDFSIKIVIGAEITRDNDGLAMSSRN 214
Db 165 ETIGGLQFAGV-PDV-----KAPLVCDFSSSILSAPLDVSKFGLIYAGA 208

QY 215 VHLSP-----EEREKALSINKSLLRAKSAAGDQV-----HCEKLTNLYIQSVT 258
Db 209 KNIGPAGLTIVIRDDLLDQAKAEIPSIKLYADQAKNGSMVNTPTSTYAWYLSGLVFEWLL 268

QY 259 DAGGRIDYAEIVDQNNLEKVE 279
Db 269 EQGG----VDAIHKNLEKAQ 285

Search completed: December 15, 2003, 13:41:33
Job time : 22 secs